

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

WO9507706 A1 9517



BASIC

95131177

J04 Z

B04 D16 S03

INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

Goodman patent

PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

A61K 38/00, C07K 5/00, C12P 21/06, C12Q 1/00, G01N 33/53	A1	(11) International Publication Number: WO 95/07706
		(43) International Publication Date: 23 March 1995 (23.03.95)
(21) International Application Number: PCT/US94/10151	(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(22) International Filing Date: 13 September 1994 (13.09.94)	Published With international search report.	
(30) Priority Data: 08/121,713 13-September 1993 (13.09.93) US	S3E4H4	
(71) Applicant: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA (US/US); 22nd floor, 300 Lakeside Drive, Oakland, CA 94612-3550 (US).		
(72) Inventors: GOODMAN, Corey, S.; UC Berkeley Campus, 519 Life Science Annex, Berkeley, CA 94720 (US). KOLODKIN, Alex, L.; UC Berkeley Campus, 519 Life Science Annex, Berkeley, CA 94720 (US). MATTHES, David; UC Berkeley Campus, 519 Life Science Annex, Berkeley, CA 94720 (US). BENTLEY, David, R.; UC Berkeley Campus, 519 Life Science Annex, Berkeley, CA 94720 (US). O'CONNOR, Timothy; UC Berkeley Campus, 519 Life Science Annex, Berkeley, CA 94720 (US).		
(74) Agents: OSMAN, Richard, Aron et al; Flehr, Hobbach, Test, Albritton & Herbert, Suite 3400, 4 Embarcadero Center, San Francisco, CA 94111-4187 (US).		
(54) Title: THE SEMAPHORIN GENE FAMILY		
(57) Abstract		
<p>A novel class of proteins, semaphorins, nucleic acids encoding semaphorins, semaphorin peptides, and methods of using semaphorins and semaphorin-encoding nucleic acids are disclosed. Semaphorin peptides and receptor agonists and antagonists provide potent modulators of nerve cell growth and regeneration. The invention provides pharmaceutical compositions, methods for screening chemical libraries for regulators of cell growth/differentiation; semaphorin gene-derived nucleic acids for use in genetic mapping, as probes for related genes, and as diagnostic reagents for genetic neurological disease; specific cellular and animal systems for the development of neurological disease therapy.</p>		

3659

95131177

THE SEMAPHORIN GENE FAMILY

The research carried out in the subject application was supported in part by grants from the National Institutes of Health. The government may have rights in any patent issuing on this application.

INTRODUCTION

Technical Field

The technical field of this invention concerns peptides, polypeptides, and polynucleotides involved in nerve cell growth.

Background

The specificity of the wiring of the nervous system — the complex pattern of specific synaptic connections — begins to unfold during development as the growing tips of neurons — the growth cones — traverse long distances to find their correct targets. Along their journey, they are confronted by and correctly navigate a series of choice points in a remarkably unerring way to ultimately contact and recognize their correct target.

The identification of growth cone guidance cues is to a large extent, the holy grail of neurobiology. These are the compounds that tell neurons when to grow, where to grow, and when to stop growing. The medical applications of such compounds and their antagonists are enormous and include modulating neuronal growth regenerative capacity, treating neurodegenerative disease, and mapping (e.g. diagnosing) genetic neurological defects.

Over decades of concentrated research, various hypotheses of chemoattractants and repellants, labeled pathways, cell adhesion molecules, etc. have been

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT:

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MY	Malawi
BB	Barbados	GR	Greece	NE	Niger
BE	Belgium	HN	Honduras	NL	Netherlands
BG	Bulgaria	IE	Ireland	NO	Norway
BR	Brazil	IT	Italy	NZ	New Zealand
CA	Canada	JP	Japan	PL	Poland
CH	Switzerland	KE	Kenya	PT	Portugal
CL	Chile	KG	Kyrgyzstan	RO	Romania
CM	Cameroun	KZ	Republic of Korea	RU	Russian Federation
CN	China	LA	Laos	SE	Sweden
CO	Colombia	LI	Liechtenstein	SI	Slovenia
CZ	Czech Republic	LK	Sri Lanka	SK	Slovakia
DE	Germany	LU	Luxembourg	SN	Senegal
DK	Denmark	LV	Latvia	TD	Chad
EE	Estonia	MC	Monaco	TG	Togo
FI	Finland	MD	Republic of Moldova	TT	Trinidad and Tobago
FR	France	MG	Madagascar	UA	Ukraine
GA	Gabon	ML	Mali	US	United States of America
		MN	Mongolia	UZ	Uzbekistan
				VN	Viet Nam

evoked to explain guidance. Recently, several recent lines of experiments suggest repulsion may play an important role in neuron guidance and two apparently unrelated factors ("Neurite Growth Inhibitor" and "Collapsin") capable of inhibiting or collapsing growth cones have been reported.

5

Relevant Literature

For a recent review of much of the literature in this field, see Goodman and Shatz (1993) *Cell* 72/Neuron 10, 77-98. A description of grasshopper fasciclin IV (now called G-Semaphorin I) appears in Kolodkin et al. (1992) *Neuron* 9, 831-845.

10 Recent reports on Collapsin and Neurite Growth Inhibitor include Raper and Kapfhammer (1990) *Neuron* 4, 21-29, an abstract presented by Raper at the GIBCO-BRL Symposium on "Genes and Development/Function of Brain" on July 26, 1993 and Schwab and Caroni (1988) *J Neurosci* 8, 2381 and Schnell and Schwab (1990) *Nature* 343, 269, respectively.

15

SUMMARY OF THE INVENTION

A novel class of proteins, semaphorins, nucleic acids encoding semaphorins, and methods of using semaphorins and semaphorin-encoding nucleic acids are disclosed. Semaphorins include the first known family of human proteins which function as growth cone inhibitors and a family of proteins involved in viral, particularly pox viral, pathogenesis and oncogenesis. Families of semaphorin-specific receptors, including receptors found on nerve growth cones and immune cells are also disclosed.

The invention provides agents, including semaphorin peptides, which specifically bind semaphorin receptors and agents, including semaphorin receptor peptides, which specifically bind semaphorins. These agents provide potent modulators of nerve cell growth, immune responsiveness and viral pathogenesis and find use in the treatment and diagnosis of neurological disease and neuro-regeneration, immune modulation including hypersensitivity and graft-rejection, 30 and diagnosis and treatment of viral and oncological infection/diseases.

Semaphorins, semaphorin receptors, semaphorin-encoding nucleic acids, and unique portions thereof also find use variously in screening chemical libraries for regulators of semaphorin or semaphorin receptor-mediated cell activity, in

2

genetic mapping, as probes for related genes, as diagnostic reagents for genetic neurological, immunological and oncological disease and in the production of specific cellular and animal systems for the development of neurological, immunological, oncological and viral disease therapy.

5

DESCRIPTION OF SPECIFIC EMBODIMENTS

The present invention discloses novel families of proteins important in nerve and immune cell function: the semaphorins and the semaphorin receptors. The invention provides agents, including semaphorin peptides, which specifically bind semaphorin receptors and agents, including semaphorin receptor peptides, which specifically bind semaphorins. These agents find a wide variety of clinical, therapeutic and research uses, especially agents which modulate nerve and/or immune cell function by specifically mimicking or interfering with semaphorin-receptor binding. For example, selected semaphorin peptides shown to act as semaphorin receptor antagonists are effective by competitively inhibiting native semaphorin association with cellular receptors. Thus, depending on the targeted receptor, these agents can be used to block semaphorin mediated neural cell growth cone repulsion or contact inhibition. Such agents find broad clinical application where nerve cell growth is indicated, e.g. traumatic injury to nerve cells, neurodegenerative disease, etc. A wide variety of semaphorin- and semaphorin receptor-specific binding agents and methods for identifying, making and using the same are described below.

Binding agents of particular interest are semaphorin peptides which specifically bind and antagonize a semaphorin receptor and semaphorin receptor peptides which specifically bind a semaphorin and prevent binding to a native receptor. While exemplified primarily with semaphorin peptides, much of the following description applies analogously to semaphorin receptor peptides.

The semaphorin peptides of the invention comprise a unique portion of a semaphorin and have semaphorin binding specificity. A "unique portion" of a semaphorin has an amino acid sequence unique to that disclosed in that it is not found in any previously known protein. Thus a unique portion has an amino acid sequence length at least long enough to define a novel peptide. Unique semaphorin portions are found to vary from about 5 to about 25 residues,

3

3 6 6 3

3 6 6 2

preferably from 5 to 10 residues in length, depending on the particular amino acid sequence. Unique semaphorin portions are readily identified by comparing the subject semaphorin portion sequences with known peptide/protein sequence data bases. Preferred unique portions derive from the semaphorin domains (which exclude the Ig-like, intracellular and transmembrane domains as well as the signal sequences) of the disclosed semaphorin sequences, especially regions that bind the semaphorin receptor, especially that of the human varieties. Preferred semaphorin receptor unique portions derive from the semaphorin binding domains, especially regions with residues which contact the semaphorin ligand, especially that of the human varieties. Particular preferred peptides are further described herein.

The subject peptides may be free or coupled to other atoms or molecules. Frequently the peptides are present as a portion of a larger polypeptide comprising the subject peptide where the remainder of the polypeptide need not be semaphorin- or semaphorin receptor-derived. Alternatively, the subject peptide may be present as a portion of a "substantially full-length" semaphorin domain or semaphorin receptor sequence which comprises or encodes at least about 200, preferably at least about 250, more preferably at least about 300 amino acids of a disclosed semaphorin/receptor sequence. Thus the invention also provides polypeptides comprising a sequence substantially similar to that of a substantially full-length semaphorin domain or a semaphorin receptor. "Substantially similar" sequences share at least about 40%, more preferably at least about 60%, and most preferably at least about 80% sequence identity. Where the sequences diverge, the differences are generally point insertions/deletions or conservative substitutions, i.e. a cysteine/threonine or serine substitution, an acidic/acidic or hydrophobic/hydrophobic amino acid substitution, an acidic/acidic or hydrophobic/hydrophobic amino acid substitution, etc.

The subject semaphorin peptides/polypeptides are "isolated", meaning unaccompanied by at least some of the material with which they are associated in their natural state. Generally, an isolated peptide/polypeptide constitutes at least about 1%, preferably at least about 10%, and more preferably at least about 50% by weight of the total peptide/protein in a given sample. By pure peptide/polypeptide is intended at least about 90%, preferably at least 95%, and more preferably at least about 99% by weight of total peptide/protein. Included in the subject peptide/polypeptide weight are any atoms, molecules, groups, or

polymers covalently coupled to the subject semaphorin/receptor peptide/polypeptide, especially peptides, proteins, detectable labels, glycosylations, phosphorylations, etc.

The subject peptides/polypeptides may be isolated or purified in a variety of ways known to those skilled in the art depending on what other components are present in the sample and to what, if anything, the peptide/polypeptide is covalently linked. Purification methods include electrophoretic, molecular, immunological and chromatographic techniques, especially affinity chromatography and RP-HPLC in the case of peptides. For general guidance in suitable purification techniques, see Scopes, R., Protein Purification, Springer-Verlag, NY (1982).

The subject peptides/polypeptides generally comprise naturally occurring amino acids but D-amino acids or amino acid mimetics coupled by peptide bonds or peptide bond mimetics may also be used. Amino acid mimetics are other than naturally occurring amino acids that conformationally mimic the amino acid for the purpose of the requisite semaphorin/receptor binding specificity. Suitable mimetics are known to those of ordinary skill in the art and include β - γ - δ amino and imino acids, cyclohexylalanine, adamantylacetic acid, etc., modifications of the amide nitrogen, the α -carbon, amide carbonyl, backbone modifications, etc. See, generally, Morgan and Gainer (1989) Ann. Repts. Med. Chem 24, 243-252; Spatola (1983) Chemistry and Biochemistry of Amino Acids, Peptides and Proteins, Vol VII (Weinstein) and Cho et al (1993) Science 261, 1303-1305 for the synthesis and screening of oligocarbamates.

The subject semaphorin peptides/polypeptides have a "semaphorin binding specificity" meaning that the subject peptide/polypeptide retains a molecular conformation specific to one or more of the disclosed semaphorins and specifically recognizable by a semaphorin-specific receptor, antibody, etc. As such, a semaphorin binding specificity may be provided by a semaphorin-specific immunological epitope, lectin binding site, etc., and preferably, a receptor binding site. Analogously, the semaphorin receptor peptides/polypeptides have a "semaphorin receptor binding specificity" meaning that these peptides/polypeptides retain a molecular conformation specific to one or more of the disclosed semaphorin receptors and specifically recognizable by a semaphorin, a receptor-specific antibody, etc.

"Specific binding" is empirically determined by contacting, for example a semaphorin-derived peptide with a mixture of components and identifying those components that preferentially bind the semaphorin. Specific binding is most conveniently shown by competition with labeled ligand using recombinant semaphorin peptide either in vitro or in cellular expression systems as disclosed herein. Generally, specific binding of the subject semaphorin has binding affinity of 10^4 M, preferably 10^5 M, more preferably 10^{10} M, under in vitro conditions as exemplified below.

The peptides/polypeptides may be modified or joined to other compounds using physical, chemical, and molecular techniques disclosed or cited herein or otherwise known to those skilled in the relevant art to affect their semaphorin binding specificity or other properties such as solubility, membrane transportability, stability, binding specificity and affinity, chemical reactivity, toxicity, bioavailability, localization, detectability, in vivo half-life, etc. as assayed by methods disclosed herein or otherwise known to those of ordinary skill in the art. For example, point mutations are introduced by site directed mutagenesis of nucleotides in the DNA encoding the disclosed semaphorin polypeptides or in the course of in vitro peptide synthesis.

Other modifications to further modulate binding specificity/affinity include chemical/enzymatic intervention (e.g. fatty acid-acylation, proteolysis,

glycosylation) and especially where the peptide/polypeptide is integrated into a larger polypeptide, selection of a particular expression host, etc. In particular, many of the disclosed semaphorin peptides contain serine and threonine residues which are phosphorylated or dephosphorylated. See e.g. methods disclosed in Roberts et al. (1991) Science 253, 1022-1026 and in Wegner et al. (1992) Science 256, 370-373. Amino and/or carboxyl termini may be functionalized e.g., for the amino group, acylation or alkylation, and for the carboxyl group, esterification or amidification, or the like. Many of the disclosed semaphorin peptides/polypeptides also contain glycosylation sites and patterns which may be disrupted or modified, e.g. by enzymes like glycosidases or used to purify/identify the receptor, e.g. with lectins. For instance, N or O-linked glycosylation sites of the disclosed semaphorin peptides may be deleted or substituted for by another basic amino acid such as Lys or His for N-linked glycosylation alterations, or deletions or polar

substitutions are introduced at Ser and Thr residues for modulating O-linked glycosylation. Glycosylation variants are also produced by selecting appropriate host cells, e.g. yeast, insect, or various mammalian cells, or by in vitro methods such as neuraminidase digestion. Useful expression systems include COS-7, 293, BHK, CHO, TM4, CV1, VERO-76, HELA, MDCK, BRL 3A, W138, Hep G2, MMT 060562, TRI cells, baculovirus systems, for examples. Other covalent modifications of the disclosed semaphorin peptides/polypeptides may be introduced by reacting the targeted amino acid residues with an organic derivatizing (e.g. methyl-3-[(p-azido-phenyl)diibio] propionimide) or crosslinking agent (e.g. 1,1-bis(diazocetyl)-2-phenylethane) capable of reacting with selected side chains or termini. For therapeutic and diagnostic localization, semaphorins and peptides thereof may be labeled directly (radioisotopes, fluorescent, etc.) or indirectly with an agent capable of providing a detectable signal, for example, a heart muscle kinase labeling site.

The following are 14 classes of preferred semaphorin peptides where bracketed positions may be occupied by any one of the residues contained in the brackets and "X" signifies that the position may be occupied by any one of the 20 naturally encoded amino acids. These enumerated peptides maintain highly conserved structures which provide important semaphorin binding specificities;

- (a) [DE]C[QKRAN]N(YFV)I (SEQ ID NO:01)
C[QKRAN]N(YFV)I[RKQT] (SEQ ID NO:02)
- (b) CGT[NG][ASN][YFHC][XRNQ] (SEQ ID NO:03)
CGT[NG][ASN]XXP (SEQ ID NO:04)
CGT[NG]XXXPX[CD] (SEQ ID NO:05)
CGTXXXPX[CD]XX[YI] (SEQ ID NO:06)
- (c) [RIQV][GA][LVK][CS]P[FY][DN] (SEQ ID NO:07)
[CS]P[FY][DN]P[DERK][HLD] (SEQ ID NO:08)
GX[GA]X[CS]PY[DN]P (SEQ ID NO:09)
- (d) L[FY]S[GA]T[VNA]A (SEQ ID NO:10)
L[FY]SXTXA[DE][FY] (SEQ ID NO:11)

[FY]S[GA]T[VNA]A[DE][FY] (SEQ ID NO:12)

(e) L[ND][AK]PNEV (SEQ ID NO:13)

5 (f) FFFRE (SEQ ID NO:14)

FF[FY]RE[TN] (SEQ ID NO:15)

10 FFRE[TN]A (SEQ ID NO:16)

F[FY]RE[TN]A (SEQ ID NO:17)

YPP[FY]RE (SEQ ID NO:18)

15 [FY]FF[FY]RE (SEQ ID NO:19)

[FY][FY][FY]RE[TN]A (SEQ ID NO:20)

[IV][FY]F[FY][FY]RE (SEQ ID NO:21)

20 D[KFY]V[FY][FYIL][FYIL][FY] (SEQ ID NO:22)

[VI][FY][FYIL][FYIL]F[RT]X[TN] (SEQ ID NO:23)

25 [VI][FY][FYIL][FYIL][FY][RT][EDV][TN] (SEQ ID NO:24)

(g) E[FY]IN[CS]CK (SEQ ID NO:25)

30 [FY]INCGK[AVI] (SEQ ID NO:26)

(h) R[VI][AG][RQ][VI]CK (SEQ ID NO:27)

R[VI]X[RQ][VI]CXXD (SEQ ID NO:28)

35 GK[VAI]XXXR[VAI]XXXCK (SEQ ID NO:29)

(i) [RKN]W[TAS][TAS][FYL]L[KR] (SEQ ID NO:30)

[FY]L[KR][AS]RL[NI]C (SEQ ID NO:31)

40 [NI]CS[IV][PS]G (SEQ ID NO:32)

W[TAS][TAS][FYL]LK[ASVIL]XL (SEQ ID NO:33)

45 W[TAS][TAS]XLKXXLXC (SEQ ID NO:34)

WX[TS]XLKXXLXC (SEQ ID NO:35)

(j) [FY][FY][ND]EIQS (SEQ ID NO:36)

50 [FY]P[FY][FY][FY][ND]E (SEQ ID NO:37)

(k) GSA[VIL]CX[FY] (SEQ ID NO:38)

55 SA[VIL]CX[FY]XM (SEQ ID NO:39)

(l) NS[NA]WL[PA]V (SEQ ID NO:40)

(m) [VLI]P[EDYSF]PRPG (SEQ ID NO:41)

[VLI]EXP[RA]PGXC (SEQ ID NO:42)

5 P[EDYSF]PRPG[TQS]C (SEQ ID NO:43)

(n) DP[HFY]C[AG]W (SEQ ID NO:44)

P[HFY]C[AG]WD (SEQ ID NO:45)

10 DPXC[AG]WD (SEQ ID NO:46)

CXXXDPXCXWD (SEQ ID NO:47)

15 CXXXDPXCXWD (SEQ ID NO:48)

CXXDPXCXWD (SEQ ID NO:49)

CXXCXXXDXXCXWD (SEQ ID NO:50)

20 CXXCXXXDXXCXWD (SEQ ID NO:51)

CXXCXXXDXXCXWD (SEQ ID NO:52)

25 The following peptides represent particularly preferred members of each class:

(a) DCQNYI (subset of SEQ ID NO:01)

30 (b) CGT[NG][AS]XXP (subset of SEQ ID NO:04)

(c) GX[SC]PYDP (subset of SEQ ID NO:09)

(d) LYSGT[VNA]A (subset of SEQ ID NO:10)

35 (e) LNAFNV (subset of SEQ ID NO:13)

(f) [FY]FF[FY]RE (SEQ ID NO:19)

(g) E[FY]IN[CS]CK (SEQ ID NO:25)

40 (h) R[VI]ARVCK (SEQ ID NO:27)

(i) W[TA][TS][FY]LK[AS]RL (subset of SEQ ID NO:33)

45 (j) PFYF[ND]EIQS (subset of SEQ ID NO:36)

(k) GSAVCX[FY] (subset of SEQ ID NO:38)

(l) NSWL[PA]V (subset of SEQ ID NO:40)

50 (m) P[ED]PRPG[TQS]C (subset of SEQ ID NO:43)

(n) DPYC[AG]WD (subset of SEQ ID NO:46)

The following 14 classes are preferred peptides which exclude semaphorin peptides encoded in open reading frames of Variola major or Vaccinia viruses.

- (a) [DE]C[QKRAN]N[YFV]I (SEQ ID NO:01)
 5 C[QKRAN]N[YFV]I[RKQT] (SEQ ID NO:02)
 (b) CGT[NG][AS][YFHC][KRHNQ] (SEQ ID NO:03)
 10 CGT[NG][ASN][YFH][KRHNQ] (SEQ ID NO:03)
 CGT[NG][AS]XNP (SEQ ID NO:04)
 (c) [RIQV][GA][LVK][CS]P[FV][DN] (SEQ ID NO:07)
 15 [CS]P[FV][DN]P[DERK][HLD] (SEQ ID NO:08)
 GX[GA]X[CS]PY[DN]P (SEQ ID NO:09)
 (d) L[FV]S[GA]T[VNA]A (SEQ ID NO:10)
 20 L[FV]SXTXA[DE][FY] (SEQ ID NO:11)
 [FY]S[GA]T[VNA]A[DE][FY] (SEQ ID NO:12)
 25 (e) L[ND][AK]PNFV (SEQ ID NO:13)
 (f) FFFRE (SEQ ID NO:14)
 30 FF[FY]RE[TN] (SEQ ID NO:15)
 FFRE[TN]A (SEQ ID NO:16)
 F[FY]RE[TN]A (SEQ ID NO:17)
 35 YFF[FY]RE (SEQ ID NO:18)
 [FY]FF[FY]RE (SEQ ID NO:19)
 [FY][FY][FY]RE[TN]A (SEQ ID NO:20)
 40 [IV][FY]F[FY][FY]RE (SEQ ID NO:21)
 D[KFY]V[FY][FYL][FYIL][FY] (SEQ ID NO:22)
 45 D[KFY]V[FY][FYIL][FYI][FY] (SEQ ID NO:22)
 [VI][FY][FYIL][FYIL]F[RT]X[TN] (SEQ ID NO:23)
 50 [VI][FY][FYIL][FYI]F[RT]X[TN] (SEQ ID NO:23)
 [VI][FY][FYIL][FYIL]FRX[TN] (SEQ ID NO:23)
 [VI][FY][FYIL][FYIL][FY][RT][EDV][TN] (SEQ ID NO:24)
 55 (g) E[FY]IN[CS]GK (SEQ ID NO:25)
- (h) R[VI][AG][RQ][VI]CK (SEQ ID NO:27)
 5 R[VI]X[RQ][VI]CXND (SEQ ID NO:28)
 GK[VAI]XXXR[VAI]XXXCK (SEQ ID NO:29)
 10 [RKN]W[TA][TAS][FYL]L[KR] (SEQ ID NO:30)
 [FY]L[KR][AS]RL[NI]C (SEQ ID NO:31)
 [NI]CS[IV][PS]G (SEQ ID NO:32)
 15 W[TA][TAS][FYL]LK[ASVIL]XL (SEQ ID NO:33)
 W[TAS][TAS][FYL]LK[ASIL]XL (SEQ ID NO:34)
 W[TA][TAS]XLXXLXC (SEQ ID NO:35)
 20 [FY][FY][ND]EIQS (SEQ ID NO:36)
 [FY]P[FY][FY][FY][ND]E (SEQ ID NO:37)
 25 (k) GSA[VIL]CX[FY] (SEQ ID NO:38)
 SA[VI]CX[FY]XM (SEQ ID NO:39)
 (l) NS[NA]WL[PA]V (SEQ ID NO:40)
 30 (m) [VLI]P[EDYSF]PRPG (SEQ ID NO:41)
 [VLI]PXPRPGXC (SEQ ID NO:42)
 35 P[EDYSF]PRPG[TQS]C (SEQ ID NO:43)
 (n) DP[HFY]C[AG]W (SEQ ID NO:44)
 P[HFY]C[AG]WD (SEQ ID NO:45)
 40 DPXC[AG]WD (SEQ ID NO:46)
 CXXXXDPXCXWD (SEQ ID NO:47)
 45 CXXDXDPXCXWD (SEQ ID NO:48)
 CXDXDPXCXWD (SEQ ID NO:49)
 CXXCXXXXXXCXWD (SEQ ID NO:50)
 50 CXXCXXXXXXCXWD (SEQ ID NO:51)
 CXXCXXXXXXCXWD (SEQ ID NO:52)

W[TAS][TAS](FYL)LK(ASVIL)XL (SEQ ID NO:33)
W[TAS][TAS]XLKXXLXC (SEQ ID NO:34)

\$ WX(TS)XLKXXLXC (SEQ ID NO:35)

(K) SA[VIL]CX{FY}XM (SEQ ID NO:39)

(III) [VLI]PXP[RA]PGXC (SEQ ID NO:42)

The disclosed semaphorin sequence data are used to define a wide variety of other semaphorin- and semaphorin receptor-specific binding agents using immunologic, chromatographic or synthetic methods available to those skilled in the art.

Of particular significance are peptides comprising unique portions of semaphorin-specific receptors and polypeptides comprising a sequence substantially similar to that of a substantially full-length semaphorin receptor. Using

semaphorin peptides, these receptors are identified by a variety of techniques known to those skilled in the art where a ligand to the target receptor is known, including expression cloning as set out in the exemplification below. For other examples of receptor isolation with known ligand using expression cloning, see, 20

Staunton et al (1989) *Nature* 339, 61; Davis et al (1991) *Science* 253, 59; Lin et al (1992) *Cell* 68, 775; Geating et al (1989) *EMBO* 8, 3667; Aruffo and Seed (1987) *PNAS* 84, 8573 and references therein. Generally, COS cells are transfected to express a cDNA library or PCR product and cells producing peptides/polypeptides which bind a semaphorin/receptor peptide/polypeptide are isolated. For

neurosemaphorin receptors, fetal brain cDNA libraries are preferred; for immunosemaphorin receptors, libraries derived from activated lymphoid or myeloid cell lines or tissue derived from sites of inflammation or delayed-type

30 hypersensitivity are preferred; and for semaphorin and semaphorin receptor

variants used by tumor cells to evade immune surveillance or suppress an immune response (oncosuppressors), libraries derived from cancerous tissue or tumor cell lines resistant to the host immune system are preferred. Alternatively, PCR primers based upon known semaphorin/receptor sequences such as those disclosed herein are used to amplify PCR product from such tissues/cells. Other

13

3673

The following 2 classes are preferred peptides which exclude semaphorin peptides encoded in open reading frames of Variola major or Vaccinia viruses

(f) YFF[FY]RE (SEQ ID NO:14)

D[RY]V[FY][FYL][FYIL][FY] (SEQ ID NO:22)

***** (SEQ ID NO:22)

.....(EVIL)F(RT)Y(TN) (SEO ID NO:23)

.....,ENVIRONMENT (SEO ID NO:23)

SECRET (SEC ID NO: 23)

15

$\lambda \in [0, 1]$

[illegible]
$$V[\mathbf{F}\mathbf{Y}][\mathbf{F}\mathbf{Y}\mathbf{T}\mathbf{L}][\mathbf{F}\mathbf{T}\mathbf{T}\mathbf{L}][\mathbf{F}\mathbf{T}\mathbf{L}]\mathbf{K}[\mathbf{L}\mathbf{B}\mathbf{V}]\mathbf{L}$$

(n) CXXXDPXCXWD (SEQ ID NO: 48)

CXXDPXCXWD (SEQ ID NO:49)

CXXCXXDXXCXWD (SEQ ID NO:51)

CXXCXXDXCXXWD (SEQ ID NO:52)

The following 5 classes are peptides which encompass peptides encoded in open reading frames of Variola major or Vaccinia viruses. Accordingly, in the event that these viral peptides are not novel per se, the present invention discloses a hitherto unforeseen and unforseeable utility for these peptides as immunosuppressants and targets of anti-viral therapy.

35 (b) CGT(NG)[ASN][YFHG][KRHNQ] (SEQ ID NO:03)

SECRET//NOFORN (SEQ ID NO: 04)

COMMUNICATIONS (SEQ ID NO:05)

40
----- (SEE ID NO: 06)

(SEC ID NO:22)

(T) [unclear] (CFO TO NO. 22)

[v1][v2][v3][v4][v5][v6][v7][v8][v9][v10][v11][v12][v13][v14][v15][v16][v17][v18][v19][v20][v21][v22][v23][v24][v25][v26][v27][v28][v29][v30][v31][v32][v33][v34][v35][v36][v37][v38][v39][v40][v41][v42][v43][v44][v45][v46][v47][v48][v49][v50][v51][v52][v53][v54][v55][v56][v57][v58][v59][v60][v61][v62][v63][v64][v65][v66][v67][v68][v69][v70][v71][v72][v73][v74][v75][v76][v77][v78][v79][v80][v81][v82][v83][v84][v85][v86][v87][v88][v89][v90][v91][v92][v93][v94][v95][v96][v97][v98][v99][v100]

$$v_{\{F1\}}[\{F11B\},\{F11B\}]=2.0$$

12

3872

receptor/ligand isolation methods using immobilized ligand or antibody are known to those skilled in the art.

Semaphorin receptor peptides with receptor binding specificity are identified by a variety of ways including having conserved consensus sequences with other semaphorin receptors, by crosslinking to ligand or receptor-specific antibody, or preferably, by screening such peptides for semaphorin binding or disruption of semaphorin-receptor binding. Methods for identifying semaphorin receptor peptides with the requisite binding activity are described herein or otherwise known to those skilled in the art. By analogous methods, semaphorin receptor peptides are used to define additional semaphorin peptides with semaphorin binding specificity, particularly receptor specificity.

The various semaphorin and semaphorin receptor peptides are used to define functional domains of semaphorins, identify compounds that associate with semaphorins, design compounds capable of modulating semaphorin-mediated nerve and immune cell function, and define additional semaphorin and semaphorin receptor-specific binding agents. For example, semaphorin mutants, including deletion mutants are generated from the disclosed semaphorin sequences and used to identify regions important for specific protein-ligand or protein-protein interactions, for example, by assaying for the ability to mediate repulsion or preclude aggregation in cell-based assays as described herein. Further, x-ray crystallographic data of the disclosed protein are used to rationally design binding molecules of determined structure or complementarity for modulating growth cone growth and guidance.

Additional semaphorin- and receptor-specific agents include specific antibodies that can be modified to a monovalent form, such as Fab, Fab', or Fv, specifically binding oligopeptides or oligonucleotides and most preferably, small molecular weight organic receptor antagonists. For example, the disclosed semaphorin and receptor peptides are used as immunogens to generate semaphorin- and receptor-specific polyclonal or monoclonal antibodies. See, Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Laboratory, for general methods. Anti-idiotypic antibody, especially internal imaging anti-ids are also prepared using the disclosures herein.

In addition to semaphorin and semaphorin-receptor derived polypeptides and peptides, other prospective agents are screened from large libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of saccharide, peptide, and nucleic acid based compounds.

Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily producible. Additionally, natural and synthetically produced libraries and compounds are readily modified through conventional chemical, physical, and biochemical means. See, e.g. Houghten et al. and Lam et al (1991) *Nature* 354, 84 and 81, respectively and Blake and Lizi-
Davis (1992), *Bioconjugate Chem* 3, 510.

Useful agents are identified with a range of assays employing a compound comprising the subject peptides or encoding nucleic acids. A wide variety of in vitro, cell-free binding assays, especially assays for specific binding to immobilized compounds comprising semaphorin or semaphorin receptor peptide find convenient use. While less preferred, cell-based assays may be used to determine specific effects of prospective agents on semaphorin-receptor binding may be assayed, see, e.g. Schnell and Schwab (1990) supra. Optionally, the intracellular C-terminal domain is substituted with a sequence encoding a oligopeptide or polypeptide domain that provides a detectable intracellular signal upon ligand binding different from the natural receptor. Useful intracellular domains include those of the human insulin receptor and the TCR, especially domains with kinase activity and domains capable of triggering calcium influx which is conveniently detected by fluorimetry by preloading the host cells with Fura-2. More preferred assays involve simple cell-free in vitro binding of candidate agents to immobilized semaphorin or receptor peptides, or vice versa. See, e.g. Fodor et al (1991) *Science* 251, 767 for light directed parallel synthesis method. Such assays are amenable to scale-up, high throughput usage suitable for volume drug screening.

Useful agents are typically those that bind to a semaphorin or disrupt the association of a semaphorin with its receptor. Preferred agents are semaphorin-specific and do not cross react with other neural or lymphoid cell membrane proteins. Useful agents may be found within numerous chemical classes, though typically they are organic compounds; preferably small organic compounds. Small organic compounds have a molecular weight of more than 150 yet less than about

4,500, preferably less than about 1500, more preferably, less than about 500.

Exemplary classes include peptides, saccharides, steroids, heterocyclics, polycyclics, substituted aromatic compounds, and the like.

Selected agents may be modified to enhance efficacy, stability, pharmaceutical compatibility, and the like. Structural identification of an agent may be used to identify, generate, or screen additional agents. For example, where peptide agents are identified, they may be modified in a variety of ways as described above, e.g. to enhance their proteolytic stability. Other methods of stabilization may include encapsulation, for example, in liposomes, etc.

The subject binding agents may be prepared in a variety of ways known to those skilled in the art. For example, peptides under about 60 amino acids can be readily synthesized today using conventional commercially available automatic synthesizers. Alternatively, DNA sequences may be prepared encoding the desired peptide and inserted into an appropriate expression vector for expression in a prokaryotic or eukaryotic host. A wide variety of expression vectors are available today and may be used in conventional ways for transformation of a competent host for expression and isolation. If desired, the open reading frame encoding the desired peptide may be joined to a signal sequence for secretion, so as to permit isolation from the culture medium. Methods for preparing the desired sequence, inserting the sequence into an expression vector, transforming a competent host, and growing the host in culture for production of the product may be found in U.S. Patent Nos. 4,710,473, 4,711,843 and 4,713,339.

For therapeutic uses, the compositions and agents disclosed herein may be administered by any convenient way. Small organics are preferably administered orally; large molecular weight (e.g. greater than 1 kD, usually greater than 3 kD, more usually greater than 10 kD) compositions and agents are preferably administered parenterally, conveniently in a pharmaceutically or physiologically acceptable carrier, e.g., phosphate buffered saline, saline, deionized water, or the like. Typically, the compositions are added to a retained physiological fluid such as blood or synovial fluid. For CNS administration, a variety of techniques are available for promoting transfer of the therapeutic across the blood brain barrier including disruption by surgery or injection, drugs which transiently open

adhesion contact between CNS vasculature endothelial cells, and compounds which facilitate translocation through such cells.

As examples, many of the disclosed therapeutics are amenable to directly injected or infused, topical, intratracheal/nasal administration, e.g. through aerosol, intraocularly, or within/on implants e.g. fibers (e.g. collagen) osmotic pumps, grafts comprising appropriately transformed cells, etc. A particularly useful application involves coating, imbedding or derivatizing fibers, such as collagen fibers, protein polymers, etc. with therapeutic peptides. Other useful approaches are described in Otto et al. (1989) *J Neuroscience Research* 22, 83-91 and Otto and

Unsicker (1990) *J Neuroscience* 10, 1912-1921. Generally, the amount

administered will be empirically determined, typically in the range of about 10 to 1000 $\mu\text{g/kg}$ of the recipient. For peptide agents, the concentration will generally be in the range of about 50 to 500 $\mu\text{g/ml}$ in the dose administered. Other additives may be included, such as stabilizers, bactericides, etc. These additives will be present in conventional amounts.

The invention provides isolated nucleic acid sequences encoding the disclosed semaphorin and semaphorin receptor peptides and polypeptides, including sequences substantially identical to sequences encoding such polypeptides. An "isolated" nucleic acid sequence is present as other than a naturally occurring chromosome or transcript in its natural state and typically is removed from at least some of the nucleotide sequences with which it is normally associated with on a natural chromosome. A complementary sequence hybridizes to a unique portion of the disclosed semaphorin sequence under low stringency conditions, for example, at 50°C and SSC (0.9 M saline/0.09 M sodium citrate) and that remains bound when subject to washing at 55°C with SSC. Regions of non-identity of complementary nucleic acids are preferably or in the case of homologous nucleic acids, a nucleotide change providing a redundant codon. A partially pure nucleotide sequence constitutes at least about 5%, preferably at least about 30%, and more preferably at least about 90% by weight of total nucleic acid present in a given fraction.

Unique portions of the disclosed nucleic acid sequence are of length sufficient to distinguish previously known nucleic acid sequences. Thus, a unique portion has a nucleotide sequence at least long enough to define a novel

oligonucleotide. Preferred nucleic acid portions encode a unique semaphorin peptide. The nucleic acids of the invention and portions thereof, other than those used as PCR primers, are usually at least about 60 bp and usually less than about 60 kb in length. PCR primers are generally between about 15 and 100 nucleotides in length.

Nucleotide (cDNA) sequences encoding several full length semaphorins are disclosed in Figs. 1-8. The invention also provides for the disclosed sequences modified by transitions, transversions, deletions, insertions, or other modifications such as alternative splicing and also provides for genomic semaphorin sequences, and gene flanking sequences, including regulatory sequences; included are DNA and RNA sequences, sense and antisense. Preferred DNA sequence portions include portions encoding the preferred amino acid sequence portions disclosed above. For antisense applications where the inhibition of semaphorin expression is indicated, especially useful oligonucleotides are between about 10 and 30

nucleotides in length and include sequences surrounding the disclosed ATG start site, especially the oligonucleotides defined by the disclosed sequence beginning about 5 nucleotides before the start site and ending about 10 nucleotides after the disclosed start site. Other especially useful semaphorin mutants involve deletion or substitution modifications of the disclosed cytoplasmic C-termini of transmembrane semaphorins. Accordingly, semaphorin mutants with semaphorin binding affinities but with altered intracellular signal transduction capacities are produced.

For modified semaphorin-encoding sequences or related sequences encoding proteins with semaphorin-like functions, there will generally be substantial sequence identity between at least a segment thereof and a segment encoding at least a portion of the disclosed semaphorin sequence, preferably at least about 60%, more preferably at least 80%, most preferably at least 90% identity. Homologous segments are particularly within semaphorin domain-encoding regions and regions encoding protein domains involved in protein-protein, particularly semaphorin-receptor interactions and differences within such segments are particularly conservative substitutions.

Typically, the invention's semaphorin peptide encoding polynucleotides are associated with heterologous sequences. Examples of such heterologous sequences include regulatory sequences such as promoters, enhancers, response elements,

signal sequences, polyadenylation sequences, etc., introns, 5' and 3' noncoding regions, etc. Other useful heterologous sequences are known to those skilled in the art or otherwise disclosed references cited herein. According to a particular embodiment of the invention, portions of the semaphorin encoding sequence are spliced with heterologous sequences to produce soluble, secreted fusion proteins, using appropriate signal sequences and optionally, a fusion partner such as β -Gal.

The disclosed sequences are also used to identify and isolate other natural semaphorins and analogs. In particular, the disclosed nucleic acid sequences are used as hybridization probes under low-stringency or PCR primers, e.g. oligonucleotides encoding functional semaphorin domains are 32 P-labeled and used to screen λ cDNA libraries at low stringency to identify similar cDNAs that encode proteins with related functional domains. Additionally, nucleic acids encoding at least a portion of the disclosed semaphorin are used to characterize tissue specific expression of semaphorin as well as changes of expression over time, particularly during organismal development or cellular differentiation.

The semaphorin encoding nucleic acids can be subject to alternative purification, synthesis, modification, sequencing, expression, transfection, administration or other use by methods disclosed in standard manuals such as Molecular Cloning, A Laboratory Manual (2nd Ed., Sambrook, Fritsch and Maniatis, Cold Spring Harbor), Current Protocols in Molecular Biology (Eds. Ausubel, Brent, Kingston, More, Feidman, Smith and Stuhl, Greene Publ. Assoc., Wiley-Interscience, NY, NY, 1992) or that are otherwise known in the art. For example, the nucleic acids can be modified to alter stability, solubility, binding affinity and specificity, etc. semaphorin-encoding sequences can be selectively

modified with a label capable of providing a detectable signal, either directly or indirectly. Exemplary labels include radioisotopes, fluorescent, biotinylation, etc. The invention also provides vectors comprising nucleic acids encoding semaphorin peptides, polypeptides or analogs. A large number of vectors, including plasmid and viral vectors, have been described for expression in a variety of eukaryotic and prokaryotic hosts. Advantageously, vectors may also include a promoter operably linked to the semaphorin-encoding portion. Vectors will often include one or more replication systems for cloning or expression, one or more

methyated, etc. The nucleic acid sequences of the present invention may also be modified with a label capable of providing a detectable signal, either directly or indirectly. Exemplary labels include radioisotopes, fluorescent, biotinylation, etc. The invention also provides vectors comprising nucleic acids encoding semaphorin peptides, polypeptides or analogs. A large number of vectors, including plasmid and viral vectors, have been described for expression in a variety of eukaryotic and prokaryotic hosts. Advantageously, vectors may also include a promoter operably linked to the semaphorin-encoding portion. Vectors will often include one or more replication systems for cloning or expression, one or more

markers for selection in the host, e.g. antibiotic resistance. The inserted semaphorin coding sequences may be synthesized, isolated from natural sources, prepared as hybrids, etc. Suitable host cells may be transformed/transfected/infected by any suitable method including electroporation, CaCl₂ mediated DNA uptake, viral infection, microinjection, microprojectile, or other methods.

Appropriate host cells include bacteria, archbacteria, fungi, especially yeast, and plant and animal cells, especially mammalian cells. Of particular interest are *E. coli*, *B. subtilis*, *Saccharomyces cerevisiae*, SF9 cells, C129 cells, 293 cells, Neurospora, and CHO, COS, HeLa cells, immortalized mammalian myeloid and lymphoid cell lines, and pluripotent cells, especially mammalian ES cells and zygotes. Preferred replication systems include M13, ColE1, SV40, baculovirus, lambda, adenovirus, AAV, BPV, etc. A large number of transcription initiation and termination regulatory regions have been isolated and shown to be effective in the transcription and translation of heterologous proteins in the various hosts. Examples of these regions, methods of isolation, manner of manipulation, etc. are known in the art. Under appropriate expression conditions, host cells can be used as a source of recombinantly produced semaphorins or analogs.

For the production of stably transformed cells and transgenic animals, nucleic acids encoding the disclosed semaphorins may be integrated into a host genome by recombination events. For example, such a sequence can be microinjected into a cell, and thereby effect homologous recombination at the site of an endogenous gene, an analog or pseudogene thereof, or a sequence with substantial identity to an semaphorin-encoding gene. Other recombination-based methods such as nonhomologous recombinations, deletion of endogenous gene by homologous recombination, especially in pluripotent cells, etc., provide additional applications. Preferred transgenics and stable transformants over-express the disclosed receptor gene and find use in drug development and as a disease model. Alternatively, knock-out cells and animals find use in development and functional studies. Methods for making transgenic animals, usually rodents, from ES cells or zygotes are known to those skilled in the art.

The compositions and methods disclosed herein may be used to effect gene therapy. See, e.g. Zhu et al. (1993) Science 261, 209-211; Gutierrez et al. (1992) Lancet 339, 715-721. For example, cells are transfected with semaphorin sequences operably linked to gene regulatory sequences capable of effecting altered semaphorin expression or regulation. To modulate semaphorin translation, cells may be transfected with complementary antisense polynucleotides. For gene therapy involving the transfection of semaphorin transfected cells, administration will depend on a number of variables that are ascertained empirically. For example, the number of cells will vary depending on the stability of the transfused cells. Transfection media is typically a buffered saline solution or other pharmacologically acceptable solution. Similarly the amount of other administered compositions, e.g. transfected nucleic acid, protein, etc., will depend on the manner of administration, purpose of the therapy, and the like.

The following examples are offered by way of illustration and not by way of limitation.

EXAMPLES

1. Isolation and characterization of Grasshopper Semaphorin I (SEQ ID NOs:57 and 58) (previously referred to as Fasciclin IV)
- In order to identify cell surface molecules that function in selective fasciculation, a series of monoclonal antibody (MAb) screens was conducted. The immunogen used for most of these screens was membranes from the longitudinal connectives (the collection of longitudinal axons) between adjacent segmental ganglia of the nervous system of the larval grasshopper. From these screens, MAb 3B11 and 8C6 were used to purify and characterize two surface glycoproteins, fasciclin I and fasciclin II, see, Bastiani et al., 1987; the genes encoding both were subsequently cloned, see, Snow et al. 1989, Zinn et al. 1988, and Harrelson and Goodman, 1988.

Another MAb isolated during these screens, MAb 6F8, was chosen for the present study because, just as with fasciclin I and fasciclin II, the antigen recognized by this MAb is expressed on a different but overlapping subset of axon pathways in the developing CNS. The 6F8 antigen appears to be localized on the outside of cell surfaces, as indicated by MAb binding when incubated both in live

preparations, and in fixed preparations in which no detergents have been added. Because the 6F8 antigen is a surface glycoprotein expressed on a subset of axon fascicles (see below), we call it fasciclin IV.

Fasciclin IV expression begins early in embryonic development before axonogenesis. At 29% of development, expression is seen on the surface of the midline mesectodermal cells and around 5-7 neuroblasts and associated ectodermal cells per hemisegment. This expression is reminiscent of the mesectodermal and neuroblast-associated expression observed with both fasciclin I and fasciclin II; however, in each case, the pattern resolves into a different subset of neuroblasts and associated ectodermal cells.

At 32% of development, shortly after the onset of axonogenesis in the CNS, fasciclin IV expression is seen on the surface of the axons and cell bodies of the three pairs of MP4, MP5, and MP6 midline progeny, the three U motoneurons, and on several unidentified neurons in close proximity to the U's. This is in contrast to fasciclin II, which at this stage is expressed on the MP1 and dMP2 neurons, and fasciclin I, which is expressed on the U neurons but not on any midline precursor progeny.

The expression of fasciclin IV on a subset of axon pathways is best observed around 40% of development, after the establishment of the first longitudinal and commissural axon pathways. At this stage, the protein is expressed on two longitudinal axon fascicles, a subset of commissural axon fascicles, a tract extending anteriorly along the midline, and a subset of fascicles in the segmental nerve (SN) and intersegmental nerve (ISN) roots.

Specifically, fasciclin IV is expressed on the U fascicle, a longitudinal pathway (between adjacent segmental neuromeres) pioneered in part by the U neurons, and on the A/P longitudinal fascicle (in part an extension of the U fascicle within each segmental neuromere. In addition, fasciclin IV is also expressed on a second narrower, medial, and more ventral longitudinal pathway. The U axons turn and exit the CNS as they pioneer the ISN; the U's and many other axons within the ISN express fasciclin IV. The continuation of the U fascicle posterior to the ISN junction is also fasciclin IV-positive. The specificity of fasciclin IV for distinct subsets of longitudinal pathways can be seen by comparing fasciclin IV and

fasciclin II expression in the same embryo; fasciclin IV is expressed on the U and A/P pathways whereas fasciclin II is expressed on the MP1 pathway.

The axons in the median fiber tract (MFT) also express fasciclin IV. The MFT is pioneered by the three pairs of progeny of the midline precursors MP4, MP5, and MP6. The MFT actually contains three separate fascicles. The axons of the two MP4 progeny pioneer the dorsal MFT fascicle and then bifurcate at the posterior end of the anterior commissure; whereas the axons of the two MP6 progeny pioneer the ventral MFT fascicle and then bifurcate at the anterior end of the posterior commissure. Fasciclin IV is expressed on the cell bodies of the six MP4, MP5, and MP6 neurons, and on their growth cones and axons as they extend anteriorly in the MFT and bifurcate in one of the two commissures. However, this expression is regional in that once these axons bifurcate and begin to extend laterally across the longitudinal pathways and towards the peripheral nerve roots, their expression of fasciclin IV greatly decreases. Thus, fasciclin IV is a label for the axons in the MFT and their initial bifurcations in both the anterior and posterior commissures. It appears to be expressed on other commissural fascicles as well. However, the commissural expression of fasciclin IV is distinct from the transient expression of fasciclin II along the posterior edge of the posterior commissure, or the expression of fasciclin I on several different commissural axon fascicles in both the anterior and posterior commissure (Bastiani et al., 1987; Harrelson and Goodman, 1988).

Fasciclin IV is also expressed on a subset of motor axons exiting the CNS in the SN. The SN splits into two major branches, one anterior and the other posterior, as it exits the CNS. Two large bundles of motoneuron axons in the anterior branch express fasciclin IV at high levels; one narrow bundle of motoneuron axons in the posterior branch expresses the protein at much lower levels. Fasciclin IV is also expressed on many of the axons in the ISN.

The CNS and nerve root expression patterns of fasciclin IV, fasciclin I, and fasciclin II at around 40% of embryonic development indicate that although there is some overlap in their patterns (e.g., both fasciclin IV and fasciclin I label the U axons), these three surface glycoproteins label distinct subsets of axon pathways in the developing CNS.

Fasciclin IV is expressed on epithelial bands in the developing limb bud

Fasciclin IV is expressed on the developing limb bud epithelium in circumferential bands; at 34.5% of development these bands can be localized with respect to constrictions in the epithelium that mark presumptive segment boundaries. In addition to a band just distal to the trochanter/coxa segment boundary, bands are also found in the tibia, femur, coxa, and later in development a fifth band is found in the tarsus. Fasciclin IV is also expressed in the nascent chordoidal organ in the dorsal aspect of the femur. The bands in the tibia, trochanter, and coxa completely encircle the limb. However, the femoral band is incomplete, containing a gap on the anterior epithelia of this segment.

The position of the T11 axon pathway with respect to these bands of fasciclin IV-positive epithelia suggests a potential role for fasciclin IV in guiding the T11 growth cones. First, the band of fasciclin IV expression in the trochanter, which is approximately three epithelial cell diameters in width when encountered by the T11 growth cones, is the axial location where the growth cones reorient from proximal migration to circumferential branch extension. The T11 cell, which marks the location of the turn, lies within this band, usually over the central or the proximal cell tier. Secondly, although there is a more distal fasciclin IV expressing band in the femur, where a change in T11 growth is not observed, there exists a gap in this band such that fasciclin IV expressing cells are not traversed by the T11 growth cones. The T11 axons also may encounter a fasciclin IV expressing region within the coxa, where interactions between the growth cones, the epithelial cells, and the Cx1 guidepost cells have not yet been investigated.

In addition to its expression over the surface of bands of epithelial cells, fasciclin IV protein, as visualized with MAb 6F8, is also found on the basal surface of these cells in a punctate pattern. This punctate staining is not an artifact of the HRP immunocytochemistry since fluorescent visualization of MAb 6F8 is also punctate. The non-neuronal expression of fasciclin IV is not restricted to limb buds. Circumferential epithelial bands of fasciclin IV expression are also seen on subesophageal mandibular structures and on the developing antennae.

MAb directed against fasciclin IV can alter the formation of the T11 axon pathway in the limb bud

The expression of fasciclin IV on an epithelial band at a key choice point in the formation of the T11 axon pathway led us to ask whether this protein is

involved in growth cone guidance at this location. To answer this question, we cultured embryos, or epithelial fillets (e.g., O'Connor et al., 1990), during the 5% of development necessary for normal pathway formation, either in the presence or absence of MAb 6F8 or 6F8 Fab fragments. Under the culture conditions used for these experiments, defective T11 pathways are observed in 14% of limbs (Chang et al., 1992); this defines the baseline of abnormalities observed using these conditions. For controls we used other MABs and their Fab fragments that either bind to the surfaces of these neurons and epithelial cells (MAb 3B11 against the surface protein fasciclin I) or do not (MAb 4D9 against the nuclear protein engrailed; Patel et al., 1989). To assess the impact of MAb 6F8 on T11 pathway formation, we compared the percentage of aberrant pathways observed following treatment with MAb 6F8 to that observed with MABs 3B11 and 4D9. Our cultures began at 32% of development when the T11 growth cones have not yet reached the epithelium just distal to the trochanter/coxa boundary and therefore have not encountered epithelial cells expressing fasciclin IV. Following approximately 30 hours in culture (~4% of development), embryos were fixed and immunostained with antibodies to HRP in order to visualize the T11 axons and other neurons in the limb bud. Criteria for scoring the T11 pathway, and the definition of "aberrant", are described in detail in the Experimental Procedures.

Although MAb 6F8 does not arrest pathway formation, several types of distinctive, abnormal pathways are observed. These defects generally begin where growth cones first contact the fasciclin IV expressing cells in the trochanter. Normally, the T11 neurons each have a single axon, and the axons of the two cells are fasciculated in that portion of the pathway within the trochanter. Following treatment with MAb 6F8, multiple long axon branches are observed within, and proximal to, the trochanter. Two major classes of pathways are taken by these branches; in 36% of aberrant limbs, multiple, long axon branches extend ventrally in the region distal to the Cx1 cells which contains the band of fasciclin IV expressing epithelial cells. In the ventral region of the trochanter, these branches

often independently turn proximally to contact the Cx1 cells, and thus complete the pathway in this region.

In the second major class of pathway defect, seen in 47% of aberrant limbs, axon branches leave the trochanter at abnormal, dorsal locations, and extend proximally across the trochanter/coxa boundary. These axons then veer ventrally, often contacting the Cx1 neurons. The remaining 17% of defects include defasciculation distal to the trochanter, axon branches that fail to turn proximally in the ventral trochanter and continue into the posterior compartment of the limb, and axon branches which cross the trochanter/coxa boundary and continue to extend proximally without a ventral turn.

When cultured in the presence of MAb 6F8, 43% of limbs exhibited malformed T11 pathways ($n = 381$) as compared to 11% with MAb 3B11 ($n = 230$) and 5% with MAb 4D9 ($n = 20$). These percentages are pooled from treatments with MAbs concentrated from hybridoma supernatant, IgGs isolated from these supernatants, and Fab fragments isolated from these IgG preparations (see Experimental Procedures). The frequency of malformed T11 pathways and the types of defects observed showed no significant variation regardless of the method of antibody preparation or type of antibody used. Since Fabs show similar results as IgGs, the effects of MAb 6F8 are not due to cross linking by the bivalent IgG.

In summary, following treatment with MAb 6F8, the T11 pathway typically exhibits abnormal morphology beginning just distal to the trochanter and at the site of fasciclin IV expression. The two most common types of T11 pathway defects described above occur in 36% of experimental limbs (treated with MAb 6F8), but are seen in only 4% of control limbs (treated with MAbs 3 B11 and 4D9).

Fasciclin IV cDNAs encode a novel integral membrane protein

Grasshopper fasciclin IV was purified by passing crude embryonic grasshopper lysates over a MAb 6F8 column. After affinity purification, the protein was eluted, precipitated, denatured, modified at cysteines, and digested with either trypsin or Lys-C. Individual peptides were resolved by reverse phase HPLC and microsequenced using standard methods.

The amino acid sequences derived from these proteolytic fragments were used to generate oligonucleotide probes for PCR experiments, resulting in products

that were used to isolate cDNA clones from the Zinn embryonic grasshopper cDNA library (Snow et al., 1988). Sequence analysis of these cDNAs reveals a single open reading frame (ORF) encoding a protein with two potential hydrophobic stretches of amino acids: an amino-terminal signal sequence of 20 residues and (beginning at amino acid 627) a potential transmembrane domain of 25 amino acids. Thus, the deduced protein has an extracellular domain of 605 amino acids, a transmembrane domain, and a cytoplasmic domain of 78 amino acids. The calculated molecular mass of the mature fasciclin IV protein is 80 kd and is confirmed by Western blot analysis of the affinity purified and endogenous protein as described below. The extracellular domain of the protein includes 16 cysteine residues that fall into three loose clusters but do not constitute a repeated domain and are not similar to other known motifs with cysteine repeats. There are also six potential sites for N-linked glycosylation in the extracellular domain.

Treatment of affinity purified fasciclin IV with N-Glycanase demonstrates that fasciclin IV does indeed contain N-linked oligosaccharides. Fasciclin IV shows no sequence similarity when compared with other proteins in the PIR data base using BLASTP (Altschul et al., 1990), and is therefore a novel type I integral membrane protein.

A polyclonal antiserum directed against the cytoplasmic domain of the protein encoded by the fasciclin IV cDNA was used to stain grasshopper embryos at 40% of development. The observed staining pattern was identical to that seen with MAb 6F8. On Western blots, this antiserum recognizes the protein we affinity purified using MAb 6F8 and then subjected to microsequence analysis. Additionally, the polyclonal serum recognizes a protein of similar molecular mass from grasshopper embryonic membranes. Taken together these data indicate that the sequence we have obtained is indeed fasciclin IV.

Four other cell surface proteins that label subsets of axon pathways in the insect nervous system (fasciclin I, fasciclin II, fasciclin III, and neuroglian) are capable of mediating homophilic cell adhesion when transfected into S2 cells in vitro (Snow et al., 1989; Elkins et al., 1990b; Grenningloh et al., 1990). To ask whether fasciclin IV can function as a homophilic cell adhesion molecule, the fasciclin IV cDNA with the complete ORF was placed under the control of the inducible metallothionein promoter (Bunch et al., 1988), transfected into S2 cells,

and assayed for its ability to promote adhesion in normally non-adhesive S2 cells. Following induction with copper, fasciclin IV was synthesized in these S2 cells as shown by Western blot analysis and cell surface staining of induced S2 cells with the polyclonal antiserum described above.

- 5 We observed no evidence for aggregation upon induction of fasciclin IV expression, thus suggesting that, in contrast to the other four proteins, fasciclin IV does not function as a homophilic cell adhesion molecule. Alternatively, fasciclin IV-mediated aggregation might require some further posttranslational modification, or co-factor, not supplied by the S2 cells, but clearly this protein acts differently in the S2 cell assay than the other four axonal glycoproteins previously tested. This is consistent with the pattern of fasciclin IV expression in the embryonic limb since only the epithelial cells and not the T11 growth cones express fasciclin IV, and yet antibody blocking experiments indicate that fasciclin IV functions in the epithelial guidance of these growth cones. Such results suggest that fasciclin IV functions in
- 10 a heterophilic adhesion or signaling system.

Discussion

Fasciclin IV is expressed on groups of axons that fasciculate in the CNS, suggesting that, much like other insect axonal glycoproteins, it functions as a homophilic cell adhesion molecule binding these axons together. Yet, in the limb bud, fasciclin IV is expressed on a band of epithelium but not on the growth cones that reorient along this band, suggesting a heterophilic function. That fasciclin IV functions in a heterophilic rather than homophilic fashion is supported by the lack of homophilic adhesion in S2 cell aggregation assays. In contrast, fasciclin I, fasciclin II, fasciclin III, and neuroglian all can function as homophilic cell adhesion molecules (Snow et al., 1989; Elkins et al., 1990b; Grenningloh et al., 1990).

cdNA sequence analysis indicates that fasciclin IV is an integral membrane protein with a novel sequence not related to any protein in the present data base. Thus, fasciclin IV represents a new type of protein that functions in the epithelial guidance of pioneer growth cones in the developing limb bud. Given its expression on a subset of axon pathways in the developing CNS, fasciclin IV functions in the guidance of CNS growth cones as well.

The results from the MAb blocking experiments illuminate several issues in T11 growth cone guidance and axon morphogenesis in the limb. First, the most striking change in growth cone behavior in the limb is the cessation of proximal growth and initiation of circumferential extension of processes upon encountering the trochanter/coxa boundary region (Bentley and Caudy, 1983; Caudy and Bentley, 1987). This could be because the band of epithelial cells within the trochanter promotes circumferential growth, or because the cells comprising the trochanter/coxa boundary and the region just proximal to it are non-permissive or aversive for growth cone migration, or both. The extension of many axon branches across the trochanter/coxa boundary following treatment with MAb 6F8 suggests that the trochanter/coxa boundary cells, which do not express fasciclin IV, are not aversive or non-permissive. Thus the change in behavior at the boundary appears to be due to the ability of fasciclin IV expressing epithelial cells to promote circumferential extension of processes from the T11 growth cones.

- 15 Secondly, treatment with MAb 6F8 results in frequent defasciculation of the axons of the two T11 neurons, and also formation of abnormal multiple axon branches, within the trochanter over fasciclin IV-expressing epithelial cells. Previous studies have shown that treatment with antibodies against ligands expressed on non-neural substrates (Landmesser et al., 1988), or putative competitive inhibitors of substrate ligands (Wang and Denburg, 1992) can promote defasciculation and increased axonal branching. Our results suggest that T11 axon:axon fasciculation and axon branching also are strongly influenced by interactions with substrate ligands, and that fasciclin IV appears to be a component of this interaction within the trochanter.

25 Thirdly, despite the effects of MAb 6F8 on axon branching, and on crossing the trochanter/coxa boundary, there remains a pronounced tendency for branches to grow ventrally both within the trochanter and within the distal region of the coxa. Consequently, all signals which can promote ventral migration of the growth cones have not been blocked by MAb 6F8 treatment. Antibody treatment may have a threshold effect in which ventral growth directing properties of fasciclin IV are more robust, and less incapacitated by treatment, than other features; alternatively, guidance information promoting ventral migration may be

independent of fasciclin IV. Time lapse video experiments to determine how the abnormal pathways we observe actually form can resolve these issues.

These results demonstrate that fasciclin IV functions as a guidance cue for the T11 growth cones just distal to the trochanter/coxa boundary, is required for these growth cones to stop proximal growth and spread circumferentially, and that the function of fasciclin IV in T11 pathway formation result from interactions between a receptor/ligand on the T11 growth cones and fasciclin IV on the surface of the band of epithelial cells results in changes in growth cone morphology and subsequent reorientation. Fasciclin IV appears to elicit this change in growth cone morphology and orientation via regulation of adhesion, a signal transduction function, or a combination of the two.

Experimental Procedures

Immunocytochemistry

Grasshopper embryos were obtained from a colony maintained at the U.C. Berkeley and staged by percentage of total embryonic development (Bentley et al., 1979). Embryos were dissected in PBS, fixed for 40 min in PEM-FA [0.1 M PIPES (pH 6.95), 2.0 mM EGTA, 1.0 mM MgSO₄, 3.7% formaldehyde], washed for 1 hr with three changes in PBT (1x PBS, 0.5% Triton X-100, 0.2% BSA), blocked for 30 min in PBT with 5% normal goat serum, and incubated overnight at 4°C in primary antibody. PBSap (1x PBS, 0.1% Saponin, 0.2% BSA) was used in place of PBT with MAb 8G7. Antibody dilutions were as follows: MAb 6F8 1:1, polyclonal antisera directed against a fasciclin IV bacterial fusion protein (#98-3) 1:400; MAb 8G7 1:4; MAb 8C6 1:1. The embryos were washed for one hour in PBT with three changes, blocked for 30 min, and incubated in secondary antibody for at least 2 hr at room temperature. The secondary antibodies were HRP-conjugated goat anti-mouse and anti-rat IgG (Jackson Immunoresearch Lab), and were diluted 1:300. Embryos were washed in PBT for one hour with three changes and then reacted in 0.5% diaminobenzidine (DAB) in PBT. The reaction was stopped with several washes in PBS and the embryos were cleared in a glycerol series (50%, 70%, 90%), mounted and viewed under Nomarski or bright field optics. For double-labelled preparations the first HRP reaction was done in PBT containing 0.06% NiCl₂ followed by washing, blocking, and incubation

30

3 6 9 0

overnight in the second primary antibody. The second antibody was visualized with a DAB reaction as described above. Embryos cultured in the presence of monoclonal antibodies were fixed and incubated overnight in goat anti-HRP (Jackson Immunoresearch Labs) conjugated to RTTC (Molecular Probes), washed for one hour in PBT with three changes, mounted in 90% glycerol, 2.5% DABCO (Polysciences), and viewed under epifluorescence. S2 cells were stained with polyclonal sera #98-3 diluted 1:400 and processed as described previously (Snow et al., 1989).

10 Monoclonal Antibody Blocking Experiments

In order to test for functional blocking, monoclonal antibody reagents were prepared as follows. Hybridoma supernatant was brought to 20% with H₂O-saturated NH₄SO₄, incubated in ice 1 hr, and spun at 15,000 g at 4°C for 20 min. The supernatant was brought to 56% with H₂O-saturated NH₄SO₄, incubated overnight at 4°C, spun as above. The pellet was resuspended in PBS using approximately 1/40 volume of the original hybridoma supernatant (often remaining a slurry) and dialyzed against 1x PBS overnight at 4°C with two changes. This reagent is referred to as "concentrated hybridoma supernatant." Purified IgG was obtained by using Immunosupure Plus Immobilized Protein A IgG Purification Kit (Pierce) to isolate IgG from the concentrated hybridoma supernatant. Fab fragments were obtained using the ImmunoPure Fab Preparation Kit (Pierce) from the previously isolated IgGs. For blocking experiments each reagent was diluted into freshly made supplemented RPMI culture media (O'Connor et al., 1990) and dialyzed overnight at 4°C against 10 volumes of the same culture media. Dilutions were as follows: concentrated hybridoma supernatant 1:4; purified IgG 150mg/ml; Fab 75mg/ml.

Embryos for culture experiments were carefully staged to between 31 and 32% of development. As embryos in each clutch typically differ by less than 1% of embryonic development from each other, the growth cones of the T11 neurons at the beginning of the culture period were located approximately in the mid-femur, well distal to the trochanter/coxa segment boundary. From each clutch at least two limbs were filleted and the T11 neurons labelled with the lipophilic dye DiI (Molecular Probes) as described (O'Connor et al., 1990) in order to confirm the

31

3 6 9 1

PCR Methods

DNA complementary to poly(A) + RNA from 45%-50% grasshopper embryos was prepared (Sambrook et al., 1989). PCR was performed using Perkin Elmer Taq polymerase (Saiki et al., 1988), and partially degenerate (based on grasshopper codon bias) oligonucleotides in both orientations corresponding to a portion of the protein sequence of several fasciclin IV peptides as determined by microsequencing. These oligonucleotides were designed so as not to include all of the peptide-derived DNA sequence, leaving a remaining 9-12 base pairs that could be used to confirm the correct identity of amplified products. All possible combinations of these sequences were tried. 40 cycles were performed, the parameters of each cycle as follows: 96°C for one min; a sequentially decreasing annealing temperature (2°C/cycle, starting at 65°C and ending at 55°C for remaining 35 cycles) for 1 min; and at 72°C for one min. Reaction products were cloned into the Sma site of M13 mp10 and sequenced. Two products, 1074 bp and 288 bp in length, contained DNA 3' to the oligonucleotide sequences encoded the additional amino acid sequence of the fasciclin IV peptide from which the oligonucleotides were derived. These two fragments have one end in common, and the oligonucleotides used to amplify them correspond to the amino acid sequences MYVQFEE and MDEAVPAF (fasciclin IV residue 29-386), and HTLMDEA and KNYVVRMDG (fasciclin IV residue 376-472).

cDNA Isolation and Sequence Analysis

Both PCR products were used to screen 1 X 10⁶ clones from a grasshopper embryonic cDNA library (Snow et al., 1988). 21 clones that hybridized to both fragments were recovered, and one 2600 bp clone was sequenced using the dideoxy chain termination method (Sanger et al., 1977) and Sequenase (US Biochemical Corp.). Templates were made from M13 mp10 vectors containing inserts generated by sonication of plasmid clones. One cDNA was completely sequenced on both strands using Oligonucleotides and double strand sequencing of plasmid DNA (Sambrook et al., 1989) to fill gaps. Two additional cDNAs were analyzed by double strand sequencing to obtain the 3' 402 bp of the transcript. All three cDNAs were used to construct a plasmid containing the entire transcript. The complete transcript sequence is 2860 bp in length with 452 bp of 5' and 217

precise location of the T11 growth cones. Prior to culturing, embryos were sterilized and dissected (Chang et al., 1992). The entire amnion and dorsal membrane was removed from the embryo to insure access of the reagents during culturing. Embryos were randomly divided into groups and cultured in one of the blocking reagents described above. Cultures were incubated with occasional agitation at 30°C for 30 hrs. At the end of the culture period embryos were fixed and processed for analysis as described above in immunocytochemistry.

For each culture experiment, the scoring of the T11 pathway in each limb was confirmed independently by a second observer. There was no statistically significant variation between the two observers. Limbs from MAB cultured embryos were compared to representative normal limbs from non-MAB cultured embryos and were scored as abnormal if any major deviation from the normal T11 pathway was observed. The T11 pathway was scored as abnormal for one or more of the following observed characteristics: (1) defasciculation for a minimum distance of approximately 25 mm anywhere along the pathway, (2) multiple axon branches that extended ventrally within the trochanter, (3) presence of one or more axon branches that crossed the trochanter/coxa boundary dorsal to the Cx1 cells, but then turned ventrally in the coxa and contacted the Cx1 cells, (4) the presence of axon branches that crossed the trochanter/coxa segment boundary, did not turn ventrally, but continued proximally toward the CNS, and (5) failure of ventrally extended axons within the trochanter to contact and reorient proximally to the Cx1 cells. For each MAB tested, the data are presented as a percentage of the abnormal T11 pathways observed. The raw data are presented in Table 1.

Protein Affinity Purification and Microsequencing

Grasshopper fasciclin IV was purified by passing crude embryonic grasshopper lysate (Bastiani et al., 1987) over an Affi-Gel 15 column (Bio Rad) conjugated with the monoclonal antibody 6F8. Protein was eluted with 50 mM DEA (pH 11.5), 0.1% Lauryldimethylamine oxide (Cal Bio Chem), and 1mM EDTA. Protein was then precipitated, denatured, modified at cysteines, and digested with either trypsin or Lys-C (Boehringer-Mannheim). Individual peptides were resolved by RP-HPLC and microsequenced (Applied Biosystems 4771 Microsequencer) using standard chemistry.

bp of 3' untranslated sequences containing stop codons in all reading frames. The predicted protein sequence was analyzed using the FASTDB and BLASTP programs (Intelligence). The fasciclin IV ORF unambiguously contains 10 of the 11 peptide sequences determined by microsequencing the fasciclin IV trypsin and Lys-C peptides.

Generation of Polyclonal Antibodies From Bacterial Fusion Proteins

Bacterial trpE fusion proteins were constructed using pATH (Koerner et al., 1991) vectors, three restriction fragments encoding extracellular sequences, and one fragment (770 bp HindIII/Eco RI, which includes amino acids 476-730) encoding both extracellular and intracellular sequences (designated #98-3). Fusion proteins were isolated by making an extract of purified inclusion bodies (Spindler et al., 1984), and rats were immunized with ~70mg of protein emulsified in RIBI adjuvant (Immunochem Research). Rats were injected at two week intervals and serum was collected 7 days following each injection. Sera were tested histologically on grasshopper embryos at 45% of development. Construct #98-3 showed a strong response and exhibited a staining pattern identical to that of MAb 6F8. Two of the extracellular constructs responded weakly but also showed the fasciclin IV staining pattern. All pre-immune sera failed to stain grasshopper embryos.

S2 Cell Transfections, Aggregation Assays, and Western Analysis

A restriction fragment containing the full length fasciclin IV cDNA was cloned into pRmHa-3 (Bunch et al, 1988) and co-transformed into *Drosophila* S2 cells (Schneider, 1972) with the plasmid pPC4 (Jokerst et al., 1989), which confers a-amanitin resistance. S2 cells were transformed using the Lipofectin Reagent and recommended protocol (BRL) with minor modifications. All other S2 cell manipulations are essentially as described (Snow et al., 1989), including adhesion assays. Fasciclin IV expression in transformed cell lines was induced for adhesion assays and histology by adding CuSO₄ to 0.7 mM and incubating for at least 48 hrs. Northern analysis confirmed transcription of fasciclin IV and surface-associated staining of the S2 cells with polyclonal serum #98-3 strongly suggests fasciclin IV is being transported to the cell surface. Preparation of membranes

from S2 cells and from grasshopper embryos. PAGE, and Western blot were performed as previously described (Elkins et al., 1990b) except that signal was detected using the enhanced chemiluminescence immunodetection system kit (Amersham). Amount of protein per lane in each sample loaded: fasciclin IV protein, ~5 ng; S2 cell membranes, 40 mg; grasshopper membranes 80 mg. Amounts of protein loaded were verified by Ponceau S staining of the blot prior to incubation with the antibody.

References cited in Example I

- 10 Altschul et al. (1990) *J. Mol. Biol.* 215:403-410; Bastiani et al. (1992) *Dev. Biol.*, in press; Bastiani et al. (1986) *J. Neurosci.* 6:3518-3531; Bastiani et al. (1986) *J. Neurosci.* 6:3542-3551; Bastiani et al. (1987) *Cell* 48:745-755; Bastiani et al. (1984) *J. Neurosci.* 4:2311-2328; Bentley and Caudy (1983) *Nature* 304:62-65; Bentley et al. (1979) *J. Embryol. Exp. Morph.* 54:47-74; Bentley and O'Connor (1992); Letourneau et al. (New York: Raven Press, Ltd.), pp. 265-282; Bunch et al. (1988) *Nucleic Acids Res.* 16:1043-1061; Chang et al. (1992) *Development* 114:507-519; Caudy and Bentley (1987) *Dev. Biol.* 119:454-465; Chou and Fasman (1974) *Biochemistry* 13:222-245; Elkins et al. (1990a) *Cell* 60:565-575; Elkins (1990b) *J. Cell Biol.* 110:1825-1832; Goodman et al. (1981) *J. Neurosci.* 20 1:94-102; Grenningloh et al. (1990) *Symp. Quam. Biol.* 55:327-340; Grenningloh et al. (1991) *Cell* 67:45-57; Harrelson and Goodman (1988) *Science* 242:700-708; Jacobs and Goodman (1989) *J. Neurosci.* 7:2402-2411; Jay and Keshishian (1990) *Nature* 348:548-551; Jokerst et al. (1989) *Mol. Gen. Genet.* 215:266-275; Koerner et al. (1991) *Methods Enzymol.* 194:477-490; Landmesser et al. (1988) *Dev. Biol.* 25 130:645-670; Lefcort and Bentley (1987) *Dev. Biol.* 119:466-480; Lefcort and Bentley (1989) *J. Cell Biol.* 108:1737-1749; O'Connor et al. (1990) *J. Neurosci.* 10:3935-3946; Patel et al. (1989) *Cell* 58:955-968; Patel et al. (1987) *Cell* 48:975-988; Raper et al. (1984) *J. Neurosci.* 4:2329-2345; Saiki et al. (1988) *Science* 239:487-494; Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual* 30 (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory); Sanger et al. (1977) *Proc. Natl. Acad. Sci. USA* 74:5463-5467; Schneider (1972) *J. Embryol. Exp. Morphol.* 27:353-365; Snow et al. (1989) *Cell* 59:313-323; Snow et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:5291-5295; Spindler et al. (1984) *J. Virol.*

49:132-141; Wang and Denburg (1992) *Neuron*. 8:701-714; Wang et al. (1992) *J. Cell Biol.* 118:163-176; and Zinn et al. (1988) *Cell* 53:577-587.

Genbank Accession Number:

5 The accession number for the sequence reported in this paper is L00709.

II. Isolation and characterization of Tribolium (SEQ ID NOs: 63 and 64) and Drosophila (SEQ ID NOs: 59 and 60) Semaphorin I, Drosophila Semaphorin II, (SEQ ID NOs: 61 and 62) Human Semaphorin III (SEQ ID NOs: 53 and 54) and Yaccinia Virus Semaphorin IV (SEQ ID NOs: 55 and 56) and Variola Major (smallpox) Virus Semaphorin IV (SEQ ID NOs: 65 and 66).

We used our G-Semaphorin I cDNA in standard low stringency screening methods (of both cDNA and genomic libraries) in an attempt to isolate a potential Semaphorin I homologue from *Drosophila*. We were unsuccessful in these screens. Since the sequence was novel and shared no similarity to anything else in the data base, we then attempted to see if we could identify a Semaphorin I homologue in other, more closely related insects. If possible, we would then compare these sequences to find the most conserved regions, and then to use probes (i.e., oligonucleotide primers for PCR) based on these conserved regions to find a *Drosophila* homologue.

In the process, we used the G-Semaphorin I cDNA in low stringency screens to clone Semaphorin I cDNAs from libraries made from locust *Locusta migratoria* embryonic RNA and from a cDNA embryonic library from the cricket *Acheta domestica*. We used PCR to clone genomic fragments from genomic DNA in the beetle *Tribolium*, and from the moth *Manduca*. We then used the *Tribolium* genomic DNA fragment to isolate cDNA clones and ultimately sequenced the complete ORF for the *Tribolium* cDNA.

In the meantime, we used the partial *Tribolium* and *Manduca* sequences in combination with the complete grasshopper sequence to identify conserved regions that allowed us to design primers for PCR in an attempt to clone a *Drosophila* Semaphorin I homologue. Several pairs of primers generated several different bands, which were subcloned and sequenced and several of the bands gave partial

sequences of the *Drosophila* Semaphorin homologue. One of the bands gave a partial sequence of what was clearly a different, more divergent gene, which we call D-Semaphorin II.

Based on the sequence of PCR products, we knew we had identified two different *Drosophila* genes, one of which appeared to be the Semaphorin I homologue, and the other a second related gene. The complete ORF sequence of the D-Semaphorin I homologue revealed an overall structure identical to G-Semaphorin I: a signal sequence, an extracellular domain of around 550 amino acids containing 16 cysteines, a transmembrane domain of 25 amino acids, and a cytoplasmic domain of 117 amino acids. When we had finished the sequence for D-Semaphorin II, we were able to begin to run homology searches in the data base, which revealed some of its structural features further described herein. The Semaphorin II sequence revealed a different structure: a signal sequence of 16 amino acids, a ~525 amino acid domain containing 16 cysteines, with a single immunoglobulin (Ig) domain of 66 amino acids, followed by a short unique region of 73 amino acids. There is no evidence for either a transmembrane domain or a potential phospholipid linkage in the C-terminus of this protein. Thus, it appears that the D-Semaphorin II protein is secreted from the cells that produce it. The grasshopper, *Tribolium*, and *Drosophila* Semaphorin I cDNA sequences, as well as the sequence of the D-Semaphorin II cDNA, are shown herein. In addition, we used this same technique to identify Semaphorin I genes in a moth, *Manduca sexta*, a locust, *Locusta migratoria*, and a cricket, *Acheta domestica*.

With this large family of insect Semaphorin genes, we identified a number of good stretches of the right amino acids (with the least degeneracy based on their codons) with strong homology for designing primers for PCR to look for human genes. We designed a set of oligonucleotide primers, and plated out several human cDNA libraries: a fetal brain library (Stratagene), and an adult hippocampus library. We ultimately obtained a human cDNA PCR bands of the right size that did not autoprime and thus were good candidates to be bonafide Semaphorin-like cDNAs from humans. These bands were purified, subcloned, and sequenced.

Whole-mount in situ hybridization experiments showed that D-Semaphorin I and II are expressed by different subsets of neurons in the embryonic CNS. D-Semaphorin I is expressed by certain cells along the midline as well as by other

neurons, whereas D-Semaphorin II is not expressed at the midline, but is expressed by a different subset of neurons. In addition, D-Semaphorin II is expressed by a subset of muscles prior to and during the period of innervation by specific motoneuron. On the polytene chromosomes, the D-Semaphorin I gene maps to (gene-band-chromosome) 29E1-22L and that of D-Semaphorin II to 53C9-102R. We have identified loss of function mutations in the D-Semaphorin I gene and a pair of P-element transposon insertions in the D-Semaphorin II gene which appear to cause severe phenotypes.

When we lined up the G-Semaphorin I, T-Semaphorin I, D-Semaphorin I, and D-Semaphorin II sequences and ran the sequences through a sequence data base in search of other sequences with significant similarity, we discovered a curious finding: these Semaphorins share sequence similarity with the A39R open reading frame (ORF) from Vaccinia virus and the A43R ORF from Variola Major (smallpox) virus and we discovered that the amino acids shared with the virus ORF were in the same regions where the insect proteins shared their greatest similarity. The viral ORF began with a putative signal sequence, continued for several hundred amino acids with sequence similarity to the Semaphorin genes, and then ended without any membrane linkage signal (suggesting that the protein as made by the infected cell would likely be secreted).

We reasoned that the virus semaphorins were appropriated host proteins advantageously exploited by the viruses, which would have host counterparts that most likely function in the immune system to inhibit or decrease an immune response, just as in the nervous system they appear to function by inhibiting growth cone extension. Analogous to situations where viruses are thought to encode a secreted form of a host cellular receptor, here the virus may cause the infected cell to make a lot of the secreted ligand to mimic an inhibitory signal and thus help decrease the immune response.

III. Isolation and characterization of Murine CNS Semaphorin III Receptor using Epitope Tagged Human Semaphorin III (hSIII)

mRNA was isolated from murine fetal brain tissue and used to construct a cDNA library in a mammalian expression vector, pCMX, essentially as in Davis et al. (1991) Science 253, 59.

38

3698

The transfection and screening procedure is modified from Lin et al (1992) Cell 68, 775. COS cells grown on glass slide flaskettes are transfected with pools of the cDNA clones, allowed to bind radiolabeled hSIII untruncated at the C-terminus end of the semaphorin domain. In parallel, similarly treated COS cells are allowed to bind unlabelled human semaphorin III truncated at the C-terminus end of the semaphorin domain and there joined to a 10-amino acid extension derived from the human c-myc proto-oncogene product. This modified hSIII allows the identification of hSIII receptors with the use of the tagged ligand as a bridge between the receptor and a murine monoclonal antibody which is specific for an epitope in the c-myc tag. Accordingly, after binding unlabelled hSIII the cells are exposed to the monoclonal which may be labeled directly or subsequently decorated with a secondary anti-mouse labeled antibody for enhanced signal amplification.

Cells are then fixed and screened using dark-field microscopy essentially as in Lin et al. (supra). Positive clones are identified and sequence analysis of murine CNS Semaphorin III receptor cDNA clones by the dideoxy chain termination method is used to construct full-length receptor coding sequences.

IV. Protocol for Protein-Protein H-Sema III - H-Sema III Receptor Drug Screening Assay.

A. Reagents:

- Neutralize Avidin: 20 µg/ml in PBS.
- Blocking buffer: 5% BSA, 0.5% Tween 20 in PBS; 1 hr, RT.
- Assay Buffer: 100 mM KCl, 20 mM HEPES pH 7.6, 0.25 mM EDTA, 1% glycerol, 0.5% NP-40, 50 mM BME, 1 mg/ml BSA, protease inhibitor cocktail.
- ³²P-H-Sema III 10x stock: 10^4 - 10^6 M "cold" truncated (Semaphorin domain) H-Sema III supplemented with 50,000-500,000 cpm of labeled and truncated H-Sema III (Beckman counter). Store at 4°C during screening.
- Protease inhibitor cocktail (100X): 1 mg Trypsin Inhibitor (BMB # 109894), 1 mg Aprotinin (BMB # 236624), 2.5 mg Benzamide (Sigma # B-6506), 2.5 mg Leupeptin (BMB # 1017128), 1 mg APMSF (BMB # 917575), and 0.2m M NaVO₃ (Sigma # S-6508) in 10 ml of PBS.

39

3699

illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

- H-Sema III Receptor: 10^{-4} - 10^{-6} M of biotinylated H-Sema III biotinylated receptor in PBS.

B. Preparation of assay plates:

- Coat with 120 μ l of stock N-Avidin per well at least 1 hr at 25°C or overnight at 4°C.

- Wash 2X with 200 μ l PBS.

- Block with 150 μ l of blocking buffer.

- Wash 2X with 200 μ l PBS.

C. Assay:

- Add 40 μ l assay buffer/well.

- Add 10 μ l candidate agent.

- Add 10 μ l 32 P-H-Sema III (5,000-50,000 cpm/0.1-10 pmoles/well = 10^5 - 10^7 M final concentration).

- Mix

- Incubate 1 hr. at 25°C.

- Add 40 μ l H-Sema III receptor (0.1-10 pmoles/40 μ l in assay buffer)

- Incubate 1 hr at 25°C.

- Stop the reaction by washing 4X with 200 μ l PBS.

- Add 150 μ l scintillation cocktail.

- Count in Topcount.

D. Assay controls (located on each plate):

a. Non-specific binding (no receptor added)

b. Soluble (non-biotinylated receptor) at 80% inhibition.

25 It is evident from the above results that one can use the methods and compositions disclosed herein for making and identifying diagnostic probes and therapeutic drugs. It will also be clear to one skilled in the art from a reading of this disclosure that advantage can be taken to effect alterations of semaphorin responsiveness in a host.

30 All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. Although the foregoing invention has been described in some detail by way of

SEQUENCE LISTINGS:

Sequences 53-68 show the nucleotide and deduced amino-acid sequences of human semaphorin III, vaccinia virus semaphorin IV, grasshopper semaphorin I, Drosophila semaphorin I, Drosophila semaphorin II, Tribolium semaphorin I and varicella major virus semaphorin IV.

SEQUENCE LISTING

10 (1) GENERAL INFORMATION:
 (1) APPLICANT: Goodman, Corey S.
 Kelodkin, Alex L.
 Matthee, David R.
 Bentley, David R.
 O'Connor, Timothy

15 (11) TITLE OF INVENTION: The Semaphorin Gene Family

20 (111) NUMBER OF SEQUENCES: 66
 (1v) CORRESPONDENCE ADDRESS:
 (A) ADDRESSEE: FLEHR HOBBACH TEST ALBRITTON & HERBERT
 (B) STREET: 4 Embarcadero Center, Suite 3400
 (C) CITY: San Francisco
 (D) STATE: CA
 (E) COUNTRY: USA
 (F) ZIP: 94111-4187

25 (v) COMPUTER READABLE FORM:
 (A) MEDIUM TYPE: Floppy disk
 (B) COMPUTER: IBM PC compatible
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

30 (vi) CURRENT APPLICATION DATA:
 (A) APPLICATION NUMBER: Not yet assigned
 (B) FILING DATE: 13-SEP-1994
 (C) CLASSIFICATION:

35 (vii) ATTORNEY/AGENT INFORMATION:
 (A) NAME: Osman, Richard A.
 (B) REGISTRATION NUMBER: 36,627
 (C) REFERENCE/DOCKET NUMBER: FP-58750-PC/RAO

40 (ix) TELECOMMUNICATION INFORMATION:
 (A) TELEPHONE: (415) 781-1989
 (B) TELEFAX: (415) 398-3249
 (C) TELEX: 910 277299 FHT UR

(2) INFORMATION FOR SEQ ID NO:1:

55 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

60 (11) MOLECULE TYPE: peptide
 (ix) FEATURE:

42

(A) NAME/KEY: Peptide
 (B) LOCATION: 1..6
 (D) OTHER INFORMATION: /label= SEQ01
 /note= "Xaa denotes D or E at residue #1; Q,K,R,A
 or N at residue #3; and Y,F or V at residue #5"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

10 Xaa Cys Xaa Asn Xaa Ile
 1 5

(2) INFORMATION FOR SEQ ID NO:2:

15 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(ix) FEATURE:

25 (A) NAME/KEY: Peptide
 (B) LOCATION: 1..6
 (D) OTHER INFORMATION: /label= SEQ02
 /note= "Xaa denotes Q,K,R,A or N at residue #2;
 Y,F or V at residue #4; and R,K,Q or T at residue #6"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

30 Cys Xaa Asn Xaa Ile Xaa
 1 5

(2) INFORMATION FOR SEQ ID NO:3:

40 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(ix) FEATURE:

50 (A) NAME/KEY: Peptide
 (B) LOCATION: 1..7
 (D) OTHER INFORMATION: /label= SEQ03
 /note= "Xaa denotes N or G at residue #4; A,S or N
 at residue #5; Y,F,H or Q at residue #6; and
 K,R,H,N or Q at residue #7"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

55 Cys Gly Thr Xaa Xaa Xaa Xaa
 1 5

(2) INFORMATION FOR SEQ ID NO:4:

60 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 8 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

3703

43

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..8(D) OTHER INFORMATION: /label= SEQ04
/note= "Xaa denotes N or G at residue #4, and A,S
or N at residue #5"

10 (x1) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Cys Gly Thr Xaa Xaa Xaa Xaa Pro
1 5

(2) INFORMATION FOR SEQ ID NO:5:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25 (11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..10(D) OTHER INFORMATION: /label= SEQ05
/note= "Xaa denotes N or G at residue #4, and C or
D at residue #10"

30 (x1) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Cys Gly Thr Xaa Xaa Xaa Xaa Pro Xaa Xaa
1 5 10

(2) INFORMATION FOR SEQ ID NO:6:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..13(D) OTHER INFORMATION: /label= SEQ06
/note= "Xaa denotes C or D at residue #10, and Y
or I at residue #13"

55 (x1) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Cys Gly Thr Xaa Xaa Xaa Xaa Pro Xaa Xaa Xaa Xaa
1 5 10

(2) INFORMATION FOR SEQ ID NO:7:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..7(D) OTHER INFORMATION: /label= SEQ07
/note= "Xaa denotes R,I,Q or V at residue #1; G or
A at residue #2; L,V or K at residue #3; C or S at
residue #4; P or Y at residue #6, and D or N at
residue #7"

15 (x1) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Xaa Xaa Xaa Xaa Pro Xaa Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:8:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..7(D) OTHER INFORMATION: /label= SEQ08
/note= "Xaa denotes C or S at residue #1; P or Y
at residue #3; D or N at residue #4; D,E,R or K at
residue #6; and H,L or D at residue #7"

35 (x1) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Xaa Pro Xaa Xaa Xaa Pro Xaa Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:9:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..9(D) OTHER INFORMATION: /label= SEQ09
/note= "Xaa denotes G or A at residue #3; C or S
at residue #5; and D or N at residue #8"

60 (x1) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Gly Xaa Xaa Xaa Xaa Pro Tyr Xaa Pro
1 5

(2) INFORMATION FOR SEQ ID NO:10:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids

(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

5

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..7

(D) OTHER INFORMATION: /label= SEQ10
/note= "Xaa denotes F or Y at residue #2; G or A at residue #4; and V, N or A at residue #6"

10

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Leu Xaa Ser Xaa Thr Xaa Ala
1 5

15

(2) INFORMATION FOR SEQ ID NO:11:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..9

(D) OTHER INFORMATION: /label= SEQ11
/note= "Xaa denotes F or Y at residue #2; D or E at residue #8; and F or Y at residue #9"

35

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Leu Xaa Ser Xaa Thr Xaa Ala Xaa Xaa
1 5

40

(2) INFORMATION FOR SEQ ID NO:12:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

45

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..8

(D) OTHER INFORMATION: /label= SEQ12
/note= "Xaa denotes F or Y at residue #1; G or A at residue #3; V, N or A at residue #5; D or E at residue #7; and F or Y at residue #8"

55

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Xaa Ser Xaa Thr Xaa Ala Xaa Xaa
1 5

60

(2) INFORMATION FOR SEQ ID NO:13:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

5

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..7

(D) OTHER INFORMATION: /label= SEQ13
/note= "Xaa denotes N or D at residue #2; and A or K at residue #3"

15

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Leu Xaa Xaa Pro Asn Phe Val
1 5

20

(2) INFORMATION FOR SEQ ID NO:14:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25

(11) MOLECULE TYPE: peptide

(1x) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Phe Phe Phe Arg Glu
1 5

35

(2) INFORMATION FOR SEQ ID NO:15:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

40

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..6

(D) OTHER INFORMATION: /label= SEQ15
/note= "Xaa denotes F or Y at residue #3; and T or N at residue #6"

45

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Phe Phe Xaa Arg Glu Xaa
1 5

55

(2) INFORMATION FOR SEQ ID NO:16:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

65

(11) MOLECULE TYPE: peptide
(1x) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..6
(D) OTHER INFORMATION: /label= SEQ16
/note= "Xaa denotes T or N at residue #5"
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:16:
Phe Phe Arg Glu Xaa Ala
1 5

15 (2) INFORMATION FOR SEQ ID NO:17:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: peptide
(1x) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..6
(D) OTHER INFORMATION: /label= SEQ17
/note= "Xaa denotes P or Y at residue #2, and T or N at residue #5"
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:17:
Phe Xaa Arg Glu Xaa Ala
1 5

35 (2) INFORMATION FOR SEQ ID NO:18:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: peptide
(1x) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..6
(D) OTHER INFORMATION: /label= SEQ18
/note= "Xaa denotes P or Y at residue #4"
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:18:
Tyr Phe Phe Xaa Arg Glu
1 5

60 (2) INFORMATION FOR SEQ ID NO:19:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: peptide
(1x) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..6
(D) OTHER INFORMATION: /label= SEQ19
/note= "Xaa denotes P or Y at residue #1, and F or Y at residue #4"
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:19:
Xaa Phe Phe Xaa Arg Glu
1 5

10 (2) INFORMATION FOR SEQ ID NO:20:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: peptide
(1x) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..7
(D) OTHER INFORMATION: /label= SEQ20
/note= "Xaa denotes P or Y at residue #1, F or Y at residue #2, P or Y at residue #3, and T or N at residue #6"
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:20:
Xaa Xaa Xaa Arg Glu Xaa Ala
1 5

35 (2) INFORMATION FOR SEQ ID NO:21:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: peptide
(1x) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..7
(D) OTHER INFORMATION: /label= SEQ21
/note= "Xaa denotes I or V at residue #1, F or Y at residue #2, Y or Y at residue #4, and P or Y at residue #5"
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:21:
Xaa Xaa Phe Xaa Xaa Arg Glu
1 5

60 (2) INFORMATION FOR SEQ ID NO:22:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide
(1x) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..7
(D) OTHER INFORMATION: /label= SEQ22
/note= "Xaa denotes K, P or Y at residue #2; F or Y
at residue #4; F, Y, I or L at residue #5; F, Y, I or
L at residue #6; and F or Y at residue #7"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:22:
Asp Xaa Val Xaa Xaa Xaa Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:23:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: peptide
(1x) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..8
(D) OTHER INFORMATION: /label= SEQ23
/note= "Xaa denotes V or I at residue #1; P or Y
at residue #2; F, Y, I or L at residue #3; F, Y, I or
L at residue #4; R or T at residue #6; and T or N
at residue #8"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:23:
Xaa Xaa Xaa Xaa Phe Xaa Xaa Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:24:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: peptide
(1x) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..8
(D) OTHER INFORMATION: /label= SEQ24
/note= "Xaa denotes V or I at residue #1; F or Y
at residue #2; F, Y, I or L at residue #3; F, Y, I or
L at residue #4; F or Y at residue #5; R or T at
residue #6; E, D or V at residue #7; and T or N at
residue #8"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:24:
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:25:

50

3710

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: peptide
(1x) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..7
(D) OTHER INFORMATION: /label= SEQ25
/note= "Xaa denotes F or Y at residue #2; and C or
S at residue #5"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:25:
Glu Xaa Ile Asn Xaa Gly Lys
1 5

(2) INFORMATION FOR SEQ ID NO:26:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: peptide
(1x) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..7
(D) OTHER INFORMATION: /label= SEQ26
/note= "Xaa denotes F or Y at residue #1; and A, V
or I at residue #7"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:26:
Xaa Ile Asn Cys Gly Lys Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:27:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: peptide
(1x) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..7
(D) OTHER INFORMATION: /label= SEQ27
/note= "Xaa denotes V or I at residue #2; A or G
at residue #3; R or Q at residue #4; and V or I at
residue #5"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:27:
Arg Xaa Xaa Xaa Xaa Cys Lys
1 5

51

3711

(2) INFORMATION FOR SEQ ID NO:28:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 9 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

- (A) NAME/KEY: Peptide
 (B) LOCATION: 1..9
 (D) OTHER INFORMATION: /label= SEQ28

/note= "Xaa denotes V or I at residue #2; R or Q at residue #4; and V or I at residue #5"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Arg Xaa Xaa Xaa Xaa Cys Xaa Xaa Asp
 1 5

(2) INFORMATION FOR SEQ ID NO:29:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 13 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

- (A) NAME/KEY: Peptide
 (B) LOCATION: 1..13
 (D) OTHER INFORMATION: /label= SEQ29

/note= "Xaa denotes V, A or I at residue #3; and V, A or I at residue #8"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Gly Lys Xaa Xaa Xaa Arg Xaa Xaa Xaa Xaa Cys Lys
 1 5 10

(2) INFORMATION FOR SEQ ID NO:30:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

- (A) NAME/KEY: Peptide
 (B) LOCATION: 1..7
 (D) OTHER INFORMATION: /label= SEQ30

/note= "Xaa denotes R, K or N at residue #1; T, A or S at residue #3; T, A or S at residue #4; F, Y or L at residue #5; and K or R at residue #7"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Xaa Trp Xaa Xaa Xaa Leu Xaa
 1 5

(2) INFORMATION FOR SEQ ID NO:31:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 8 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

- (A) NAME/KEY: Peptide
 (B) LOCATION: 1..8
 (D) OTHER INFORMATION: /label= SEQ31

/note= "Xaa denotes P or Y at residue #1; K or R at residue #3; A or S at residue #4; and N or I at residue #7"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Xaa Leu Xaa Xaa Arg Leu Xaa Cys
 1 5

(2) INFORMATION FOR SEQ ID NO:32:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

- (A) NAME/KEY: Peptide
 (B) LOCATION: 1..6
 (D) OTHER INFORMATION: /label= SEQ32

/note= "Xaa denotes N or I at residue #1; I or V at residue #4; and P or S at residue #5"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Xaa Cys Ser Xaa Xaa Gly
 1 5

(2) INFORMATION FOR SEQ ID NO:33:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 9 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

- (A) NAME/KEY: Peptide
 (B) LOCATION: 1..9
 (D) OTHER INFORMATION: /label= SEQ33

/note= "Xaa denotes T, A or S at residue #2; T, A or S at residue #3; F, Y or L at residue #4; and A, S, V, I or L at residue #7"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Trp Xaa Xaa Xaa Leu Lys Xaa Xaa Leu
1 5

5 (2) INFORMATION FOR SEQ ID NO:34:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION: 1..11

(D) OTHER INFORMATION: /label= SEQ34

/note= "Xaa denotes T, A or S at residue #3; and

T, A or S at residue #3"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Trp Xaa Xaa Xaa Leu Lys Xaa Xaa Leu Xaa Cys
1 5 10

30 (2) INFORMATION FOR SEQ ID NO:35:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION: 1..11

(D) OTHER INFORMATION: /label= SEQ35

/note= "Xaa denotes T or S at residue #3"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Trp Xaa Xaa Xaa Leu Lys Xaa Xaa Leu Xaa Cys
1 5 10

50 (2) INFORMATION FOR SEQ ID NO:36:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION: 1..7

(D) OTHER INFORMATION: /label= SEQ36

/note= "Xaa denotes F or Y at residue #1; F or Y

at residue #2; and N or D at residue #3"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Xaa Xaa Xaa Glu Ile Gln Ser
1 5

5 (2) INFORMATION FOR SEQ ID NO:37:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION: 1..7

(D) OTHER INFORMATION: /label= SEQ37

/note= "Xaa denotes F or Y at residue #1; F or Y

at residue #3; F or Y at residue #4; F or Y at

residue #5; and N or D at residue #6"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Xaa Pro Xaa Xaa Xaa Xaa Glu
1 5

(2) INFORMATION FOR SEQ ID NO:38:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION: 1..7

(D) OTHER INFORMATION: /label= SEQ38

/note= "Xaa denotes V, I or L at residue #4; and F

or Y at residue #7"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Gly Ser Ala Xaa Cys Xaa Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:39:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(1x) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION: 1..8

(D) OTHER INFORMATION: /label= SEQ39

/note= "Xaa denotes V, I or L at residue #3; and F

or Y at residue #6"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Ser Ala Xaa Cys Xaa Xaa Xaa Met
1 5

5

(2) INFORMATION FOR SEQ ID NO:40:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10

(11) MOLECULE TYPE: peptide

15

(1x) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION: 1..7

(D) OTHER INFORMATION: /label= SEQ40
/note= "Xaa denotes N or A at residue #3, and P or A at residue #6"

20

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Asn Ser Xaa Trp Leu Xaa Val
1 5

25

(2) INFORMATION FOR SEQ ID NO:41:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

30

(11) MOLECULE TYPE: peptide

35

(1x) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION: 1..7

(D) OTHER INFORMATION: /label= SEQ41
/note= "Xaa denotes V, L or I at residue #1, and E, D, Y, S or F at residue #3"

45

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Xaa Pro Xaa Pro Arg Pro Gly
1 5

50

(2) INFORMATION FOR SEQ ID NO:42:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

55

(11) MOLECULE TYPE: peptide

60

(1x) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION: 1..9

(D) OTHER INFORMATION: /label= SEQ42
/note= "Xaa denotes V, L or I at residue #1, and R or A at residue #5"

65

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Xaa Pro Xaa Pro Xaa Pro Gly Xaa Cys
1 5

5

(2) INFORMATION FOR SEQ ID NO:43:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10

(11) MOLECULE TYPE: peptide

15

(1x) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION: 1..8

(D) OTHER INFORMATION: /label= SEQ43
/note= "Xaa denotes E, D, Y, S or F at residue #2, and T, Q or S at residue #7"

20

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Pro Xaa Pro Arg Pro Gly Xaa Cys
1 5

25

(2) INFORMATION FOR SEQ ID NO:44:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

30

(11) MOLECULE TYPE: peptide

35

(1x) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION: 1..6

(D) OTHER INFORMATION: /label= SEQ44
/note= "Xaa denotes H, P or Y at residue #3, and A or G at residue #5"

45

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:44:

Asp Pro Xaa Cys Xaa Trp
1 5

50

(2) INFORMATION FOR SEQ ID NO:45:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

55

(11) MOLECULE TYPE: peptide

60

(1x) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION: 1..6

(D) OTHER INFORMATION: /label= SEQ45
/note= "Xaa denotes H, P or Y at residue #2, and A or G at residue #4"

65

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Pro Xaa Cys Xaa Trp Asp
1 5

5

(2) INFORMATION FOR SEQ ID NO:46:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10

(11) MOLECULE TYPE: peptide

15

(1x) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..7
(D) OTHER INFORMATION: /label= SEQ46
/note= "Xaa denotes A or G at residue #5"

20

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:46:

Asp Pro Xaa Cys Xaa Trp Asp
1 5

25

(2) INFORMATION FOR SEQ ID NO:47:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

30

(11) MOLECULE TYPE: peptide

35

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Cys Xaa Xaa Xaa Asp Pro Xaa Cys Xaa Trp Asp
1 5 10

40

(2) INFORMATION FOR SEQ ID NO:48:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

45

(11) MOLECULE TYPE: peptide

50

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Cys Xaa Xaa Xaa Asp Pro Xaa Cys Xaa Trp Asp
1 5 10

55

(2) INFORMATION FOR SEQ ID NO:49:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

60

65

(11) MOLECULE TYPE: peptide

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Cys Xaa Xaa Asp Pro Xaa Cys Xaa Trp Asp
1 5 10

5

(2) INFORMATION FOR SEQ ID NO:50:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10

15

(11) MOLECULE TYPE: peptide

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Cys Xaa Xaa Cys Xaa Xaa Xaa Asp Xaa Xaa Cys Xaa Trp Asp
1 5 10 15

20

(2) INFORMATION FOR SEQ ID NO:51:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25

30

(11) MOLECULE TYPE: peptide

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Cys Xaa Xaa Cys Xaa Xaa Xaa Asp Xaa Xaa Cys Xaa Trp Asp
1 5 10

35

40

(2) INFORMATION FOR SEQ ID NO:52:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

45

50

(11) MOLECULE TYPE: peptide

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:52:

Cys Xaa Xaa Cys Xaa Xaa Xaa Asp Xaa Xaa Cys Xaa Trp Asp
1 5 10

55

(2) INFORMATION FOR SEQ ID NO:53:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2601 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

60

65

(11) MOLECULE TYPE: cDNA

(1x) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 16..2331

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:53:

5	CGAATTCCTT	CGAC	ATG	GGC	TGG	TTA	ACT	AGG	ATT	GTC	TGT	CTT	TTC	TGG	51	TGC	ANG	AAT	GAC	TTT	GGA	GGG	CAC	AGA	AGT	CTG	CTG	AAT	AAA	TGG	ACA	867		
															10	Cys	Lys	Asn	Asp	Phe	Gly	His	Arg	Ser	Leu	Val	Asn	Lys	Trp	Thr				
															15																			
10	GGA	GTA	TTA	CTT	ACA	GCA	AGA	AAC	TAT	CAG	AAT	GGG	AAG	AAC	99	ACA	TTC	CTC	AAA	GCT	CGT	CTG	ATT	TGC	TCA	GTG	CCA	GCT	CCA	AAT	GGC	915		
															20	Thr	Phe	Leu	Lys	Ala	Arg	Leu	Ile	Cys	Ser	Val	Pro	Gly	Pro	Asn	Gly	300		
															25																			
15	GTC	CCA	AGG	CTG	AAA	TTA	TCC	TAC	AAA	GAA	ATC	TTG	GAA	TCC	147	ATC	CAC	ACT	CAT	TTT	GAT	GAA	CTG	CAG	GAT	GTA	TTT	CTA	ATG	AAC	TTT	963		
															30	Val	Pro	Arg	Leu	Lys	Leu	Ser	Tyr	Lys	Glu	Met	Leu	Glu	Met	Asn	Phe	315		
															35																			
20	GTC	ATC	ACT	ATC	ATC	ATC	GGC	TTG	GGC	AAC	AGC	TCC	ACT	TAT	195	AAA	GAT	CCT	AAA	AAT	CCA	GTT	GTA	TAT	GGA	GTG	TTT	ACG	ACT	TCC	AGT	1011		
															40	Val	Ile	Thr	Phe	Asn	Gly	Leu	Ala	Asn	Ser	Tyr	Gly	Val	Phe	Thr	Ser	Ser		
															45																			
25	CTT	TTG	GAT	CAG	GAA	CGG	AGT	AGC	CTG	TAT	GTT	GGA	GCA	AAG	243	AAC	ATT	TTT	ANG	GGA	TCA	CCC	CTG	TGT	ATG	TAT	AGC	ATG	ACT	GAT	CTG	1059		
															20	Leu	Leu	Asp	Glu	Gly	Arg	Ser	Arg	Leu	Val	Gly	Ala	Lys	Asp	His				
															25																			
30	ATA	TTT	TCA	TTC	GAC	CTG	CTT	AAT	ATC	AAG	GAT	TTT	CAA	AAG	291	AGA	AGG	GTG	TTC	CTT	CGT	CCA	TAT	CCC	CAC	AGG	GAT	GGA	CCC	AAC	TAT	1107		
															30	Ile	Phe	Ser	Phe	Asp	Leu	Val	Asn	Ile	Lys	Asp	Phe	Gln	Lys	Ile	Val			
															35																			
35	TGG	CCA	GTA	TCT	TAC	ACC	AGA	AGA	GAT	GAA	TGC	TGG	GCT	GGA	339	CMA	TGG	GTG	CTT	TAT	CAA	AGA	GTG	CCC	TAT	CCA	CGG	CCA	GGA	ACT	1155			
															40	Trp	Pro	Val	Ser	Tyr	Thr	Arg	Asp	Glu	Cys	Lys	Trp	Ala	Gly	Lys				
															45																			
40	GAC	ATC	CTG	AAA	GAA	TGT	GCT	AAT	ATC	ATC	ATC	ATC	ATC	ATC	387	TCT	CCC	AGC	AAA	ACA	TTT	GCT	GGT	TTT	GAC	TCT	ACA	AAG	GAC	CTT	CCT	1203		
															40	Asp	Ile	Leu	Lys	Glu	Cys	Ala	Asn	Phe	Ile	Lys	Val	Leu	Lys	Ala	Tyr			
															45																			
45	AAT	CAG	ACT	CAC	TTG	TAC	GGC	TGT	GGA	AGC	GGG	GCT	TTT	CAT	435	GAT	GAT	GTT	ATA	ACC	TTT	GCA	AGA	AGT	CAT	CCA	GGC	ATG	TAC	AAT	CCA	1251		
															40	Asn	Gln	Thr	His	Leu	Tyr	Ala	Cys	Gly	Thr	Gly	Ala	Phe	His	Pro	Ile			
															45																			
50	TGC	ACC	TAC	ATT	GAA	ATT	GGA	CAT	CAT	CCT	CAG	GAC	AAT	ATT	483	GTG	TTT	CCT	ATG	AAC	AAT	CGC	CCA	ATA	GTG	ATC	AAA	ACC	GAT	GTA	AAT	1299		
															50	Cys	Thr	Tyr	Ile	Gly	Ile	Gly	His	Pro	Glu	Asp	Asn	Ile	Phe	Lys				
															55																			
55	CTG	GAG	AAC	TCA	CAT	TTT	GAA	AAC	GGC	GCT	GGG	AAG	AGT	CCA	531	TAT	CAA	TTT	ACA	CAA	ATT	GTG	GTA	GAC	GTG	GAT	GCA	GAA	GAT	GCA	1347			
															40	Leu	Glu	Asn	Ser	His	Phe	Glu	Asn	Gly	Arg	Gly	Lys	Ser	Pro	Tyr	Asp			
															45																			
60	CCT	AAG	CTG	ACA	GCA	TCC	CTT	TTA	ATA	GAT	GGA	GAA	TTA	TAC	579	CAG	TAT	GAT	GTT	ATG	TTT	ATC	GGA	ACA	GAT	GTT	GGG	ACC	GTT	CTT	AAA	1395		
															50	Pro	Lys	Leu	Leu	Thr	Ala	Ser	180											
															55																			
65	GGA	ACT	GCA	GCT	GAT	TTT	ATG	GGG	GGA	GAC	TTT	GCT	ATC	TTC	627	GTA	CTT	TCA	ATT	CCT	ANG	GAG	ACT	TGG	TAT	GAT	TTA	GAA	GAG	GTT	CTG	1443		
															60	Gly	Thr	Ala	Ala	Asp	Phe	Met	Gly	Arg	Asp	Phe	Ala	Ile	Phe	Arg	Thr			
															65																			
70	CTT	GGG	CAC	CAC	CCA	ATC	AGG	ACA	GAG	CAG	CAT	GAT	TCC	AGG	675	CTG	GAA	GAA	ATG	ACA	GTT	TTT	CGG	GAA	CCG	ACT	GCT	ATT	TCA	GCA	ATG	1491		
															70	Leu	Gly	His	His	Pro	Ile	Arg	Thr	Glu	Gln	His	Asp	Ser	Arg	Trp				
															75																			
75	CTC	AAT	GAT	CCA	ANG	TTT	ATT	AGT	GGC	CAC	CTC	ATC	TCA	GAG	723	CTC	AAT	GAT	CCA	ANG	TTT	ATT	AGT	GGC	CAC	CTA	TAT	ATT	GCT	TCA	ACG	GCT	GGG	1539
															80	Leu	Asn	Asp	Pro	Lys	Phe	Ile	Ser	Ala	His	Leu	Ile	Ser	Glu	Ser	Asp			
															85																			
80	AAT	CCT	GAA	GAT	GAC	AAA	GTA	TAC	TTT	TTT	CTC	GCT	GAA	AAT	771	GTT	CCC	CAG	CTC	CCT	TTA	CAC	GGG	TGT	GAT	ATT	TAC	GGG	AAA	GGG	TGT	1587		
															80	Asn	Pro	Glu	Asp	Asp	Lys	Val	Tyr	Phe	Phe	Arg	Glu	Asn	Ala	Ile				

5 GCT GAG TGT TGC CTC GCG CCA GAT CTT TGT TGT GGT GAT CCT TCT 1635
 Ala Glu Cys Cys Leu Ala Arg Asp Pro Tyr Cys Ala Trp Asp Gly Ser 530
 5 GCA TGT TCT CGC TAT TTT CCC ACT GCA AAG ACA CGC ACA AGA CAA 1683
 Ala Cys Ser Arg Tyr Phe Pro Thr Ala 550
 10 CAT ATA AGA AAT GGA GAC CCA CTG ACT CAC TGT TCA GAC TTA CAC CAT 1731
 Asp Ile Arg Asn Gly Asp Pro Leu Thr His Cys Ser Asp Leu His His 570
 15 GAT AAT CAC CAT GCG CAC AGC CCG CCA GAG ACA ATC ATC TAT GGT CTA 1779
 Asp Asn His His Gly His Ser 580
 20 GAG AAT AGT AGC ACA TTT TTG GAA TGC AGT CCG AAG TCG CAG AGA CCG 1827
 Glu Asn Ser Ser Thr Phe Leu Glu Cys Ser Pro Lys Ser Gln Arg Ala 600
 25 CTC GTC TAT TGC CAA TTC CAG AGG CGA AAT GAA GAG CCA AAA GAA GAG 1875
 Leu Val Tyr Trp Gln Phe Gln Arg Asn 615
 30 ATC AGA CTG GAT CAT CAT ATC ATC AGC ACA GAT CAA GCG CTT CTG CTA 1923
 Ile Arg Val Asp Asp His Ile Ile Arg Thr Asp Gln Gly Leu Leu 625
 35 CGT AGT CTA CAA CAG CAG GAT TCA GGC AAT TAC CTC TGC CAT CCG GTG 1971
 Arg Ser Leu Gln Gln Lys Asp Ser Gly Asn Tyr Leu Cys His Ala Val 645
 40 GAA CAT GCG TTC ATA CAA ACT CTT CTT ANG CTA ACC CTG GAA CTC ATT 2019
 Glu His Gly Phe Ile Gln Thr 660
 45 GAC ACA CAG CAT TTG GAA GAA CTT CTT CAT AAA GAT GAT GAT CCA GAT 2067
 Asp Thr Glu His Leu Glu Glu Leu Leu His Lys Asp Asp Asp Gly Asp 680
 50 GCG TCT TAG ACC AAA GAA ATG TCC AAT AGC ATG ACA CCT AGC CAG AAG 2115
 Gly Ser Lys Thr Lys Glu Met Ser Asn Ser Met Thr Pro Ser Gln Lys 695
 55 GTC TGC TAC AGA CAC TTC ATG CAG CTC ATC AAC CAC CCC AAT CTC AAC 2163
 Val Trp Tyr Arg Asp Phe Met Gln Leu Ile Asn His Pro Asn Leu Asn 715
 60 AGC ATG CAT GAG TTC TGT GAA CAA GTT TGG AAA AGG GAG CCA AAA CAA 2211
 Thr Met Asp Glu Phe Cys Glu Gln Val Trp Lys Arg Asp Arg Lys Gln 730
 65 CGT CCG CAA AGG CCA CCA CAT ACC CCA GGG AAC AGT AAC AAA TGG AAG 2259
 Arg Arg Gln Arg Pro Gly His Thr Pro Gly Asn Ser Asn Lys Trp Lys 745
 70 CAC TTA CAA GAA AAT AAG AAA GGT AGA AAC AGG AGG ACC CAC GAA TTT 2307
 His Leu Gln Glu Asn Lys Lys Gly Arg Asn Arg Arg Thr His Glu Phe 755
 75 GAG AGG CCA CCC AGG ACT GTC TGACCTGCAT TACCTCTACA AACCTCAAC 2358
 Glu Arg Ala Pro Arg Ser Val 770
 80 AAGTAGAAC TTGCTAGAC AATACTGCA AAACAAATG CAATATACAT GAACTTTTT 2418
 CATGCCATTA TGTGATGTT TACATGCTG GGAATTCAG CTGAGTTCCA CCATATATA 2478

ATTAAATCCA TGAATTAATCTT TCTAATAGG CTTTTTTTCT CTAATACCAC CCGGTTAAAA 2538
 CTAAAGACA CCGTAACCTT CCGTACGCCA TTCATACAGG TCCTATTATTA AGCAACGGAA 2598
 5 TTC 2601
 (2) INFORMATION FOR SEQ ID NO154:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 771 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (11) MOLECULE TYPE: protein
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO154:
 1 Met Gly Trp Leu Thr Arg Ile Val Cys Leu Phe Thr Trp Gly Val Leu Leu 15
 20 Thr Ala Arg Ala Asn Tyr Gln Asn Gly Lys Asn Asn Val Pro Arg Leu 30
 25 Lys Leu Ser Tyr Lys Glu Met Leu Glu Ser Asn Asn Val Ile Thr Phe 45
 30 Asn Gly Leu Ala Asn Ser Ser Tyr His Thr Phe Leu Leu Asp Glu 60
 35 Glu Arg Ser Arg Leu Tyr Val Gly Ala Lys Asp His Ile Phe Ser Phe 75
 40 Asp Leu Val Asn Ile Lys Asp Phe Gln Lys Ile Val Trp Pro Val Ser 90
 45 Tyr Thr Arg Arg Asp Glu Cys Lys Trp Ala Gly Lys Asp Ile Leu Lys 110
 50 Glu Cys Ala Asn Phe Ile Lys Val Leu Lys Ala Tyr Asn Gln Thr His 125
 55 Leu Tyr Ala Cys Gly Thr Gly Ala Phe His Pro Ile Cys Thr Tyr Ile 140
 60 Glu Ile Gly His His Pro Glu Asp Asn Ile Phe Lys Leu Glu Asn Ser 155
 65 His Phe Glu Asn Gly Arg Gly Lys Ser Pro Tyr Asp Pro Lys Leu Leu 170
 70 Thr Ala Ser Leu Leu Ile Asp Gly Glu Leu Tyr Ser Gly Thr Ala Ala 185
 75 Asp Phe Met Gly Arg Asp Phe Ala Ile Phe Arg Thr Leu Gly His His 200
 80 His Pro Ile Arg Thr Glu Gln His Asp Ser Arg Trp Leu Asn Asp Pro 215
 85 Lys Phe Ile Ser Ala His Leu Ile Ser Glu Ser Asp Asn Pro Glu Asp 230
 90 Asp Lys Val Tyr Phe Phe Arg Glu Asn Ala Ile Asp Gly Glu His 245
 95 Ser Gly Lys Ala Thr His Ala Arg Ile Gly Gln Ile Cys Lys Asn Asp 260
 100

Phe Gly Gly His Arg Ser Leu Val Asn Lys Trp Thr Thr Phe Leu Lys
275 280

Ala Arg Leu Ile Cys Ser Val Pro Gly Pro Asn Gly Ile Asp Thr His
290 295 300

Phe Asp Glu Leu Gln Asp Val Phe Leu Met Asn Phe Lys Asp Pro Lys
305 310 315 320

Asn Pro Val Val Tyr Gly Val Phe Thr Thr Ser Ser Asn Ile Phe Lys
325 330 335

Gly Ser Ala Val Cys Met Tyr Ser Met Ser Asp Val Arg Arg Val Phe
340 345 350

Leu Gly Pro Tyr Ala His Arg Asp Gly Pro Asn Tyr Gln Trp Val Pro
355 360 365

Tyr Gln Gly Arg Val Pro Tyr Pro Arg Pro Gly Thr Cys Pro Ser Lys
370 375 380

Thr Phe Gly Gly Phe Asp Ser Thr Lys Asp Leu Pro Asp Asp Val Ile
385 390 395 400

Thr Phe Ala Arg Ser His Pro Ala Met Tyr Asn Pro Val Phe Pro Met
405 410 415

Asn Asn Arg Pro Ile Val Ile Lys Thr Asp Val Asn Tyr Gln Phe Thr
420 425 430

Gln Ile Val Val Asp Arg Val Asp Ala Glu Asp Gly Gln Tyr Asp Val
435 440 445

Met Phe Ile Gly Thr Asp Val Gly Thr Val Leu Lys Val Val Ser Ile
450 455 460

Pro Lys Glu Thr Trp Tyr Asp Leu Glu Glu Val Leu Leu Glu Met
465 470 475 480

Thr Val Phe Arg Glu Pro Thr Ala Ile Ser Ala Met Glu Leu Ser Thr
485 490 495

Lys Gln Gln Gln Leu Tyr Ile Gly Ser Thr Ala Gly Val Ala Gln Leu
500 505 510

Pro Leu His Arg Cys Asp Ile Tyr Gly Lys Ala Cys Ala Glu Cys
515 520 525

Leu Ala Arg Asp Pro Tyr Cys Ala Trp Asp Gly Ser Ala Cys Ser Arg
530 535 540

Tyr Phe Pro Thr Ala Lys Arg Arg Thr Arg Arg Gln Asp Ile Arg Asn
545 550 555 560

Gly Asp Pro Leu Thr His Cys Ser Asp Leu His His Asp Asn His
565 570 575

Gly His Ser Pro Glu Glu Arg Ile Ile Tyr Gly Val Glu Asn Ser Ser
580 585 590

Thr Phe Leu Glu Cys Ser Pro Lys Ser Gln Arg Ala Leu Val Tyr Trp
595 600 605

Gln Phe Gln Arg Arg Asn Glu Glu Arg Lys Glu Ile Arg Val Asp
610 615 620

Asp His Ile Ile Arg Thr Asp Gln Gly Leu Leu Arg Ser Leu Gln
625 630 635 640

Gln Lys Asp Ser Gly Asn Tyr Leu Cys His Ala Val Glu His Gly Phe
645 650 655

Ile Gln Thr Leu Leu Lys Val Thr Leu Glu Val Ile Asp Thr Glu His
660 665 670

Leu Glu Glu Leu Leu His Lys Asp Asp Asp Gly Asp Gly Ser Lys Thr
675 680 685

Lys Glu Met Ser Asn Ser Met Thr Pro Ser Gln Lys Val Trp Tyr Arg
690 695 700

Asp Phe Met Gln Leu Ile Asn His Pro Asn Leu Asn Thr Met Asp Glu
705 710 715 720

Phe Cys Glu Gln Val Trp Lys Arg Asp Arg Lys Gln Arg Arg Gln Arg
725 730 735

Pro Gly His Thr Pro Gly Asn Ser Asn Lys Trp Lys His Leu Gln Glu
740 745 750

Asn Lys Lys Gly Arg Asn Arg Arg Thr His Glu Phe Glu Arg Ala Pro
755 760 765

Arg Ser Val
770

(2) INFORMATION FOR SEQ ID NO:55:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1332 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(1x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 7..1329

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:55:

CGAATA ATG ATG GTA TTA TTA CAT CCT GTA TAC TCT ATA CTC TTT GTA
Met Met Val Leu Leu His Ala Val Tyr Ser Ile Val Phe Val
1 5 10

GAT GTT ATA ATC ATA AAA GTA CAG AGG TAT ATC AAC GAT ATT CTA ACT
Asp Val Ile Ile Ile Lys Val Gln Arg Tyr Ile Asn Asp Ile Leu Thr
15 20 25 30

CTT GAC ATT TTT TAT TTA TTT AAA ATG ATA CCT TTG TTA TTT ATT TTA
Leu Asp Ile Phe Tyr Leu Phe Lys Met Ile Pro Leu Leu Phe Ile Leu
35 40 45

TTC TAT TTT GCT AAC GGT ATC GAA TGG CAT AAG TTT GAA ACG AGT GAA
Phe Tyr Phe Ala Asn Gly Ile Glu Trp His Lys Phe Glu Thr Ser Glu
50 55 60

GAA ATA ATT TCT ACT TAC TTA TTA GAC GAC GAT TTA TAC ACG GGT GTT
Glu Ile Ile Ser Thr Tyr Leu Leu Asp Asp Val Leu Thr Thr Gly Val
65 70 75

AAT GCG GCG GTA TAC ACA TTT TCA AAT AAT AAA CTA AAC AAA ACT CGT
Asn Gly Ala Val Tyr Thr Thr Ser Asn Asn Lys Leu Asn Lys Thr Gly
80 85 90 288

TTA ACT AAT AAT AAT ATA ACA ACA TCT ATA AAA GTA GAG GAT GCG 336
 Leu Thr Aen Aen Aen Tyr Ile Thr Thr Ser Ile Lys Val Glu Asp Ala 110
 95 100 105 110 115

5 GAT AAG GAT ACA TTA TGC CGA ACC AAT AAC GGA AAT CCC AAA TGT 384
 Asp Lys Asp Thr Leu Val Cys Gly Thr Aen Aen Gly Aen Pro Lys Cys 125
 115 120 125 130 135

10 TGG AAA ATA GAC GGT TCA GAC GAC CCA AAA CAT AGA GGT AGA GCA TAC 432
 Trp Lys Ile Asp Gly Ser Asp Pro Lys His Arg Gly Arg Gly Tyr 140
 130 135 140 145 150

15 GCT CCT TAT CAA AAT AGC AAA GTA ACG ATA ATC AGT CAC AAC GCA TGT 480
 Ala Pro Tyr Gln Aen Ser Lys Val Thr Ile Ile Ser His Aen Gly Cys 155
 145 150 155 160 165

20 GTA CTA TCT CAC ATA AAC ATA TCA AAA GAA GCA ATT AAA CCA TGC AGA 528
 Val Leu Ser Asp Ile Aen Ile Ser Lys Glu Gly Ile Lys Arg Trp Arg 170
 160 165 170 175 180

AGA TTT GAC GGA CCA TGT GGT TAT GAT TTA TAC ACG GCG GAT AAC GTA 576
 Arg Phe Asp Gly Pro Cys Gly Tyr Asp Leu Tyr Thr Ala Asp Aen Val 185
 175 180 185 190 195

25 ATT CCA AAA GAT GGT TTA CCA GGA GCA TTC GTC CAT AAA GAT GGT ACT 624
 Ile Pro Lys Asp Gly Leu Arg Gly Ala Phe Val Asp Lys Asp Gly Thr 205
 195 200 205 210 215

30 TAT GAC AAA GTT TAC ATT CTT TTC ACT GAT ACT ATC GCG TCA AAG AGA 672
 Tyr Asp Lys Val Tyr Ile Leu Phe Thr Asp Thr Ile Gly Ser Lys Arg 220
 210 215 220 225 230

35 ATT GTC AAA ATT CCG TAT ATA GCA CAA ATG TGC GTA AAC GAC GAA GGT 720
 Ile Val Lys Ile Pro Tyr Ile Ala Gln Met Cys Leu Aen Asp Glu Gly 235
 225 230 235 240 245

GCT CCA TCA TCA TTG TCT AGT CAT AGA TGG TCG ACG TTT CTC AAA GTC 768
 Gly Pro Ser Ser Leu Ser Ser His Arg Trp Ser Thr Phe Leu Lys Val 250
 240 245 250 255 260

40 GAA TTA CAA TGT GAT ATC GAC CCA AGA AGT TAT AGA CAA ATT ATT CAT 816
 Glu Leu Glu Cys Asp Ile Asp Gly Arg Ser Tyr Arg Gln Ile Ile His 270
 255 260 265 270 275

45 TCT AGA ACT ATA AAA ACA GAT AAT GAT ACG ATA CTA TAT GTA TTC TTC 864
 Ser Arg Thr Ile Lys Thr Asp Aen Asp Thr Ile Leu Tyr Val Phe Phe 285
 275 280 285 290 295

50 GAT AGT CCT TAT TCC AAG TCC GCA TTA TGT ACC TAT TCT ATG AAT ACC 912
 Asp Ser Pro Tyr Ser Lys Ser Ala Leu Cys Thr Tyr Ser Met Aen Thr 300
 290 295 300 305 310

55 ATT AAA CAA TCT TTT TCT ACG TCA AAA TTG GAA GCA TAT ACA AAG CAA 960
 Ile Lys Gln Ser Phe Ser Thr Ser Lys Leu Glu Gly Tyr Thr Lys Gln 315
 305 310 315 320 325

TTG CCG TCG CCA GCC TCT GGT ATA TGT CTA CCA CCT GCA AAA GTT GTT 1008
 Leu Pro Pro Pro Ala Ser Gly Ile Cys Leu Pro Ala Gly Lys Val Val 330
 320 325 330 335 340

60 CCA CAT ACC ACG TTT GAA GTC ATA GAA AAA TAT AAT GTA CTA GAT GAT 1056
 Pro His Thr Thr Phe Phe Val Ile Glu Lys Tyr Aen Val Leu Asp Asp 350
 335 340 345 350 355

65 ATT ATA AAG CCT TTA TCT AAC CAA CCT ATC TTC GAA GCA CCG TCT GGT 1104
 Ile Ile Lys Pro Leu Ser Aen Gln Pro Ile Phe Glu Gly Pro Ser Gly 365
 355 360 365 370 375

GTT AAA TGG TTC GAT ATA AAG GAG AAG GAA AAT GAA CAT CCG GAA TAT 1152
 Val Lys Trp Phe Asp Ile Lys Glu Lys Glu Aen Glu His Arg Glu Tyr 375
 370 375 380 385 390

5 AGA ATA TAC TTC ATA AAA GAA AAT TCT ATA TAT TCG TTC GAT ACA AAA 1200
 Arg Ile Tyr Phe Ile Lys Glu Aen Ser Ile Tyr Ser Phe Asp Thr Lys 395
 385 390 395 400 405

10 TCT AAA CAA ACT CCT ACG TCG CAA GTC GAT CCG CCA CTA TTT TCA GTA 1248
 Ser Lys Gln Thr Arg Ser Ser Gln Val Asp Ala Arg Leu Phe Ser Val 410
 400 405 410 415 420

ATC GTA ACT TCG AAA CCG TTA TTT ATA GCA GAT ATA GCG ATA GGA GTA 1296
 Met Val Thr Ser Lys Pro Leu Phe Ile Ala Asp Ile Gly Ile Gly Val 425
 415 420 425 430 435

20 GGA ATG CCA CAA ATG AAA AAA ATA CTT AAA ATG TAA 1332
 Gly Met Pro Gln Met Lys Lys Ile Leu Lys Met 435
 430 435 440 445 450

(2) INFORMATION FOR SEQ ID NO156:

(1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 441 amino acids
 (B) TYPE: amino acid
 (C) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO156:

Met Met Val Leu Leu His Ala Val Tyr Ser Ile Val Phe Val Asp Val 1
 1 5 10 15 20

Ile Ile Ile Lys Val Gln Arg Tyr Ile Aen Asp Ile Leu Thr Leu Asp 25
 20 25 30 35 40

Ile Phe Tyr Leu Phe Lys Met Ile Pro Leu Leu Phe Ile Leu Phe Tyr 35
 35 40 45 50 55

Phe Ala Aen Gly Ile Glu Trp His Lys Phe Glu Thr Ser Glu Glu Ile 50
 50 55 60 65 70

Ile Ser Thr Tyr Leu Leu Asp Asp Val Leu Tyr Thr Gly Val Aen Gly 65
 65 70 75 80 85

Ala Val Tyr Thr Phe Ser Aen Aen Lys Leu Aen Lys Thr Gly Leu Thr 85
 85 90 95 100 105

Aen Aen Aen Tyr Ile Thr Thr Ser Ile Lys Val Glu Asp Ala Asp Lys 100
 100 105 110 115 120

Asp Thr Leu Val Cys Gly Thr Aen Aen Gly Aen Pro Lys Cys Trp Lys 115
 115 120 125 130 135

Ile Asp Gly Ser Asp Asp Pro Lys His Arg Gly Arg Gly Tyr Ala Pro 130
 130 135 140 145 150

Tyr Gln Aen Ser Lys Val Thr Ile Ile Ser His Aen Gly Cys Val Leu 145
 145 150 155 160 165

Ser Asp Ile Aen Ile Ser Lys Glu Gly Ile Lys Arg Trp Arg Arg Phe 165
 165 170 175 180 185

Asp Gly Pro Cys Gly Tyr Asp Leu Tyr Thr Ala Asp Aen Val Ile Pro 180
 180 185 190 195 200

Lys Asp Gly Leu Arg Gly Ala Phe Val Asp Lys Asp Gly Thr Tyr Asp
 195 200
 Lys Val Tyr Ile Leu Phe Thr Asp Thr Ile Gly Ser Lys Arg Ile Val
 210 215
 Lys Ile Pro Tyr Ile Ala Gln Met Cys Leu Asn Asp Glu Gly Gly Pro
 225 230 235
 Ser Ser Leu Ser Ser His Arg Trp Ser Thr Phe Leu Lys Val Glu Leu
 245 250

15

Glu Cys Asp Ile Asp Gly Arg Ser Tyr Arg Gln Ile Ile Glu Ser Arg	260	270
---	-----	-----

Pro Tyr Ser Lys Ser Ala Leu Cys Thr Tyr Ser Met Asn Thr Ile Lys
275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995 1000

Gln Ser Phe Ser Thr Ser Lys Leu Glu Gly Tyr Thr Lys Gln Leu Pro
305 310 315 320

25 Ser Pro Ala Ser Gly Ile Cys Leu Pro Ala Gly Lys Val Val Phe His 335
325 330

30 Lys Pro Leu Ser Asn Gln Pro Ile Phe Glu Gly Pro Ser Gly Val Lys
340 345 350
165

35 Trp Phe Asp Ile Lys Lys Glu Asn Glu His Arg Glu Tyr Arg Ile 380
370 375

Tyr Phe Ile Lys Glu Asn Ser Ile Tyr Ser Phe Asp Thr Lys Ser Val
385 390 395 400

[illegible]

45 Pro Cln Met Lys Lys Ile Leu Lys Met 440
435

50 (2) INFORMATION FOR SEQ ID NO:57:

(A) LENGTH: 2854 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(11) MOLECULE TYPE: cDNA

(11) NAME/KEY: CDS
(B) LOCATION: 451..2640

65 ATTCGACGTC CGCGTACAG CCTACGCCGC GACGATCTTT CCTCTGGCCA GCGGAAACT 60

89

Phe Ile Phe Phe Phe Arg Glu Thr Ala Val Glu Tyr Ile Asn Cys 220 225 230
 CGA AAG GCT ATC TAT TCA ACA GTT GCC AGA CTC TGT AAA CAT GAC AAG 1194
 5 Cys Lys Ala Ile Tyr Ser Arg Val Ala Arg Val Cys Lys His Asn Lys 235 240 245
 GGC GGC CCT CAT CAG GGT GGT GAC AGA TGG ACT TCT TTT TTG AAA TCA 1242
 10 Gly Gly Pro His Gln Gly Gly Asp Arg Trp Thr Ser Phe Leu Lys Ser 250 255 260
 CGT CTG AAC TGT TGC GTC CCT GCA CAT TAT CCA TTT TAC TTC AAT GAA 1290
 Arg Leu Asn Cys Ser Val Pro Gly Asp Tyr Pro Phe Tyr Phe Asn Glu 265 270 275
 15 ATT CAG TCA ACA AGT GAC ATC ATT CAA GCA AAT TAT GGT GGT CAA GTC 1338
 Ile Gln Ser Thr Ser Asp Ile Ile Glu Gly Asn Tyr Gly Gly Gln Val 280 285 290
 20 GAG AAA CTC ATC TAC GGT TTC AGC ACA CCA CTC AAC TCT ATT GGT 1386
 Glu Lys Leu Ile Tyr Gly Val Phe Thr Thr Pro Val Asn Ser Ile Gly 300 305 310
 GGC TCT GCT GTT TGT GCC TTC ACT ATG AAG TCA ATA CTT GAG TCA TTT 1434
 25 Gly Ser Ala Val Cys Ala Phe Ser Met Lys Ser Ile Leu Glu Ser Phe 315 320 325
 CAT GGT CCA TTT AAA GAG CAG GAA AGC ATG AAC TCA AAC TGG TTG CCA 1482
 Asp Gly Pro Phe Lys Glu Gln Glu Thr Met Asn Ser Asn Trp Leu Ala 330 335 340
 30 GTG CCA AGC CTT AAA GTG CCA GAA CCA AGG CCT GCA CAA TGT GTG AAT 1530
 Val Pro Ser Leu Lys Val Pro Glu Pro Arg Pro Gly Gln Cys Val Asn 345 350 355
 35 GAC AGT GCT ACA CTT CCT GAT GTG TCT GTC AAT TTT GTA AAG TCA CAT 1578
 Asp Ser Arg Thr Leu Pro Asp Val Ser Val Asn Phe Val Lys Ser His 360 365 370
 40 ACA CTG ATG GAT GAG GCC GTG CCA GCA TTT TTT ACT CGG CCA ATT CTC 1626
 Thr Leu Met Asp Glu Ala Val Pro Ala Phe Phe Thr Arg Pro Ile Leu 375 380 385
 45 ATT CGG ATC AGC TTA CAG TAC AGA TTT ACA AAA ATA GCT GTT GAT CAA 1674
 Ile Arg Ile Ser Leu Gln Tyr Arg Phe Thr Lys Ile Ala Val Asp Gln 390 400 405
 50 CAA GTC CCA ACA CCA GAT GGG AAA CCG TAT GAT GTC CTG TTT ATA CGA 1722
 Gln Val Arg Thr Pro Asp Gly Lys Ala Tyr Asp Val Leu Phe Ile Gly 410 415 420
 55 ACT GAT GAT GGC AAA GTG ATA AAA GCT TTG AAC TCT GCC TCC TTT GAT 1770
 Thr Asp Asp Gly Lys Val Ile Lys Ala Leu Asn Ser Ala Ser Phe Asp 425 430 435
 TCA TCT GAT ACT GTA GAT ACT GTT GTA ATA GAA GAA CTC GAA GTG TTG 1818
 Ser Ser Asp Thr Val Asp Ser Val Val Ile Glu Glu Leu Gln Val Leu 440 445 450
 60 CCA CCT GCA GTA CCT GTT AAG AAC CTG TAT GTG GTG CCA ATG GAT GCG 1866
 Pro Pro Gly Val Pro Val Lys Asn Leu Tyr Val Val Arg Met Asp Gly 455 460 465
 65 CAT GAT AGC AAG CTG GTT GTC TCT GAT GAT GAT GAT GAT GAT GAT GAT 1914
 Asp Asp Ser Lys Leu Val Val Val Ser Asp Asp Glu Ile Leu Ala Ile 470 475 480
 AAG CTT CAT GCT TGT GGC TCA GAT AAA ATA ACA AAT TGT CCA GAA TGT 1962

Lys Leu His Arg Cys Gly Ser Asp Lys Ile Thr Asn Cys Arg Glu Cys 490 495 500
 5 GTG TCC TTG CAA CAT CTT TAT TGT TCA TGG CAC GAT AAT GTA GAA TTA AAA 2010
 Val Ser Leu Gln Asn Pro Tyr Cys Ala Trp Asp Asn Val Glu Leu Lys 505 510 515
 TGT ACA GCT GTA GGT TCA CCA GAC TGG ACT GCT GCA AAA ACA GCG TTT 2058
 10 Cys Thr Ala Val Gly Ser Pro Asp Trp Ser Ala Gly Lys Arg Arg Phe 520 525 530
 ATT CAG AAC ATT TCA CTC GGT GAA CAT AAA GCT TGT GGT GGA GGT CCA 2106
 Ile Gln Asn Ile Ser Leu Gly Glu His Lys Ala Cys Gly Gly Arg Pro 535 540 545
 15 CAA ACA GAA ATC GTT GCT TCT TCT GTA CCA ACT CAG CCG ACG ACA AAA 2154
 Gln Thr Glu Ile Val Ala Ser Pro Val Pro Thr Gln Pro Thr Thr Lys 550 555 560
 20 TCT AGT GGC GAT CCC GTT CAT TCA ATC CAC CAG GCT GAA TTT GAA CCT 2202
 Ser Ser Gly Asp Pro Val His Ser Ile His Gln Ala Glu Phe Glu Pro 565 570 580
 25 GAA ATT CAC AAC CAG ATT GTT ATT CGA GTA GAT CAC AGC AAC GTC ATT 2250
 Glu Ile Asp Asn Glu Ile Val Ile Gly Val Asp Asp Ser Asn Val Ile 585 590 595
 30 CCT AAT ACC CTG GCT GAA ATA AAT CAT GCT TCA AAG CTG CCT TCC 2298
 Pro Asn Thr Leu Ala Glu Ile Asn His Ala Gly Ser Lys Leu Pro Ser 600 605 610
 TCC CAG GAA AAG TTG CCT ATT TAT ACA GCG GAG ACT CTG ACT ATT GCT 2346
 Ser Gln Glu Lys Leu Pro Ile Tyr Thr Ala Glu Thr Leu Thr Ile Ala 615 620 625
 35 ATA GTT ACA TCA TCC CTT GCA GCT CTA GTT GTT GGT TTC ATC TCT GGA 2394
 Ile Val Thr Ser Cys Leu Gly Ala Leu Val Val Gly Phe Ile Ser Gly 630 635 640
 40 TTT CTT TTT TCT CGG CCA TGC AGG GCA GAG GAT TAC ACA GAC ATG CCT 2442
 Phe Leu Phe Ser Arg Arg Cys Arg Gly Glu Asp Tyr Thr Asp Met Pro 645 650 660
 45 TTT CCA GAT CAA CCC CAT CAG CTA AAT AGG CTC ACT GAG GCT GGT CTG 2490
 Phe Pro Asp Gln Arg His Gln Leu Asn Arg Leu Thr Glu Ala Gly Leu 665 670 675
 50 AAT GCA GAC TCA CCC TAT CTT CCA CCC TGT GCC AAT AAC AAG GCA GCG 2538
 Asn Ala Asp Ser Pro Tyr Leu Pro Pro Cys Ala Asn Asn Lys Ala Ala 680 685 690
 ATA NAT CTT GTG CTC AAT GTC CCA CCA AAG AAT GCA AAT GCA AAA AAT 2586
 Ile Asn Leu Val Leu Asn Val Pro Pro Lys Asn Ala Asn Gly Lys Asn 695 700 705
 55 CCC AAC TCT TCA GCT GAA AAC AAA CCA ATA CAG AAA GTA AAA AAG ACA 2634
 Ala Asn Ser Ser Ala Glu Asn Lys Pro Ile Gln Lys Val Lys Lys Thr 710 715 720
 TAC ATT TACAGCAAT CTTTGGTATC TGTTCGGTG CACGCCCATG CCACTAGAT 2690
 Tyr Ile 725 730 735
 65 AACCAAGACT CTATTAGAA ATGCTCTCAA GAAAGTTAAA AACATCTAGA CTTCTGTAAT 2750
 CGAGACACC ACTTCCATA GTAATACAGA ACAATGTAA ATAAATCTA CAGAAGAGCT 2810

CTTCTTAC CAAAGAGT TATAGTATC TGTGATCAGT TTGC

2854

(2) INFORMATION FOR SEQ ID NO:58:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 730 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:58:

15 Met Arg Ala Ala Leu Val Ala Val Ala Leu Leu Trp Val Ala Leu 15
 1 His Ala Ala Ala Trp Val Asn Asp Val Ser Pro Lys Met Tyr Val Gln 30
 20 Phe Gly Glu Glu Arg Val Gln Arg Phe Leu Gly Asn Glu Ser His Lys 45
 25 Asp His Phe Lys Leu Leu Glu Lys Asp His Asn Ser Leu Leu Val Gly 60
 30 Thr Glu Gln Arg Ile Glu Trp His Ser Ser Gly Ala His Arg Glu Leu 95
 35 Cys Tyr Leu Lys Gly Lys Ser Glu Asp Asp Cys Gln Asn Tyr Ile Arg 110
 40 Ala Tyr Lys Pro Leu Cys Arg His Tyr Ala Leu Lys Asp Gly Asp Tyr 140
 45 Asp His Asn Ser Thr Ala Ile Tyr Ser Glu Gly Gln Leu Tyr Ser Ala 175
 50 Thr Val Ala Asp Phe Ser Gly Thr Asp Pro Leu Ile Tyr Arg Gly Pro 190
 55 Val Asn Thr Met Glu Tyr Asn Asp Phe Ile Phe Phe Phe Arg Glu 220
 60 Ala Arg Val Cys Lys His Asp Lys Gly Gly Pro His Gln Gly Gly Asp 255
 65 Arg Trp Thr Ser Phe Leu Lys Ser Arg Leu Asn Cys Ser Val Pro Gly 270
 Asp Tyr Pro Phe Tyr Phe Asn Glu Ile Gln Ser Thr Ser Asp Ile Ile 285

72

3732

Glu Gly Asn Tyr Gly Gly Gln Val Glu Lys Leu Ile Tyr Gly Val Phe 290
 5 Thr Thr Pro Val Asn Ser Ile Gly Gly Ser Ala Val Cys Ala Phe Ser 320
 Met Lys Ser Ile Leu Glu Ser Phe Asp Gly Pro Phe Lys Glu Gln Glu 335
 10 Thr Met Asn Ser Asn Trp Leu Ala Val Pro Ser Leu Lys Val Pro Glu 345
 15 Pro Arg Pro Gly Gln Cys Val Asn Asp Ser Arg Thr Leu Pro Asp Val 365
 Ser Val Asn Phe Val Lys Ser His Thr Leu Met Asp Glu Ala Val Pro 380
 20 Ala Phe Phe Thr Arg Pro Ile Leu Ile Arg Ile Ser Leu Gln Tyr Arg 395
 Phe Thr Lys Ile Ala Val Asp Gln Gln Val Arg Thr Pro Asp Gly Lys 410
 25 Ala Tyr Asp Val Leu Phe Ile Gly Thr Asp Asp Gly Lys Val Ile Lys 425
 30 Ala Leu Asn Ser Ala Ser Phe Asp Ser Ser Asp Thr Val Asp Ser Val 445
 Val Ile Glu Glu Leu Gln Val Leu Pro Pro Gly Val Pro Val Lys Asn 460
 35 Leu Tyr Val Val Arg Met Asp Gly Asp Asp Ser Lys Leu Val Val Val 475
 Ser Asp Asp Glu Ile Leu Ala Ile Lys Leu His Arg Cys Gly Ser Asp 490
 40 Lys Ile Thr Asn Cys Arg Glu Cys Val Ser Leu Gln Asp Pro Tyr Cys 505
 45 Ala Trp Asp Asn Val Glu Leu Lys Cys Thr Ala Val Gly Ser Pro Asp 525
 Trp Ser Ala Gly Lys Arg Arg Phe Ile Gln Asn Ile Ser Leu Gly Glu 540
 50 His Lys Ala Cys Gly Gly Arg Pro Gln Thr Glu Ile Val Ala Ser Pro 555
 Val Pro Thr Gln Pro Thr Thr Lys Ser Ser Gly Asp Pro Val His Ser 575
 55 Ile-His Gln Ala Glu Phe Glu Pro Glu Ile Asp Asn Glu Ile Val Ile 590
 60 Gly Val Asp Asp Ser Asn Val Ile Pro Asn Thr Leu Ala Glu Ile Asn 605
 His Ala Gly Ser Lys Leu Pro Ser Ser Gln Glu Lys Leu Pro Ile Tyr 620
 65 Thr Ala Glu Thr Leu Thr Ile Ala Ile Val Thr Ser Cys Leu Gly Ala 640

3733

73

Leu Val Val Gly Phe Ile Ser Gly Phe Leu Phe Ser Arg Arg Cys Arg
645 650 655
5 Gly Glu Asp Tyr Thr Asp Met Pro Phe Pro Asp Gln Arg His Gln Leu
660 665 670
Asn Arg Leu Thr Glu Ala Gly Leu Asn Ala Asp Ser Pro Tyr Leu Pro
675 680 685
10 Pro Cys Ala Asn Asn Lys Ala Ala Ile Asn Leu Val Leu Asn Val Pro
690 695 700
Pro Lys Asn Ala Asn Gly Lys Asn Ala Asn Ser Ser Ala Glu Asn Lys
705 710 715 720
15 Pro Ile Gln Lys Val Lys Lys Thr Tyr Ile
725 730

20 (2) INFORMATION FOR SEQ ID NO:59:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3560 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(12) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 1..1953

(13) SEQUENCE DESCRIPTION: SEQ ID NO:59:

35 GAG GAT GAT TGT CAG AAT TAC ATC CGC ATC ATG CTG CCA TCG CCG
1 Gln Asp Asp Cys Gln Asn Tyr Ile Arg Ile Met Val Val Pro Ser Pro
5 10 15
40 GGT CGC GTT TTC GTT TGT GGC ACC AAC TCG TTC CGG CCC ATG TCC AAC
20 25 30
Gly Arg Leu Phe Val Cys Gly Thr Asn Ser Phe Arg Pro Met Cys Asn
45 ACC TAT ATC ATT AGT GAC AGC AAC TAC AGC CTG GAG GCC AGC AAG AAC
35 Thr Tyr Ile Ile Ser Asp Ser Asn Tyr Thr Leu Glu Ala Thr Lys Asn
50 GCA CAG CCG GTG TCG CCC TAC GAT CCA GGT CAC AAC TCC ACC TCT GTG
60 Gly Gln Ala Val Cys Pro Tyr Asp Pro Arg His Asn Ser Thr Ser Val
CTG GCC GAC AAC GAA CTG TAT TCC GGT ACC GTG CCG GAT TTC AGT GGC
65 70 75 80
Leu Ala Asp Asn Glu Leu Tyr Ser Gly Thr Val Ala Asp Phe Ser Gly
55 ACC GAT CCG ATT ATC TAC CCG GAG CCC CTG CAG ACC GAG CAG TAC GAT
Ser Asp Pro Ile Ile Tyr Arg Glu Pro Leu Gln Thr Glu Gln Tyr Asp
85 90 95
60 ACC CTA AGT CTC AAC GCA CCG AAC TTT GTG AGC TCA TTT AGC CAG GGC
100 110 115
Ser Leu Ser Leu Asn Ala Pro Asn Phe Val Ser Ser Phe Thr Gln Gly
384 GAC TTT CTC TAT TTC TTT CCG GAA ACC GCC GTT GAG TTT ATC AAC
115 120 125
Asp Phe Val Tyr Phe Phe Arg Glu Thr Ala Val Glu Phe Ile Asn
TGT GGC AAG GCG ATT TAT TCG CCG GTT GCC CCG GTC TGC AAA TGG GAC 432

Cys Gly Lys Ala Ile Tyr Ser Arg Val Ala Arg Val Cys Lys Trp Asp
130 135 140
5 AAA GGT GGC CCG CAT CGA TTC CGC AAC CCG TCG ACA TCC TTC CTC AAG
145 150 155 160
Lys Gly Gly Pro His Arg Phe Arg Asn Arg Trp Thr Ser Phe Leu Lys
TCC CGC CTC AAC TCC TCC ATT CGC GGC GAT TAT CCT TTC TAC TTT AAT
165 170 175 180
Ser Arg Leu Asn Cys Ser Ile Pro Gly Asp Tyr Pro Phe Tyr Phe Asn
185 190 195 200
10 GAA ATC CAA TCT CCC ACC AAT CTG GTG GAG GGA CAG TAT GGC TCG ATG
Glu Ile Gln Ser Ala Ser Asn Leu Val Glu Gly Gln Tyr Gly Ser Met
205 210 215 220
15 AGC TCG AAA CTG ATC TAC GCA GTC TTC AAC CCG AGC AAC TCA ATT
Ser Ser Lys Leu Ile Tyr Gly Val Phe Asn Thr Pro Ser Asn Ser Ile
225 230 235 240
20 CGC GGC TCA CCG GTT TGT GGC TTT GGC CTC CAG GAC ATT GCC GAT ACG
Pro Gly Ser Ala Val Cys Ala Phe Ala Leu Gln Asp Ile Ala Asp Thr
245 250 255 260
25 TTT CAG GGT CAG TTC AAG CAG CAG ACT GGC ATC AAC TCC AAC TGG CTG
Phe Glu Gly Gln Phe Lys Glu Gln Thr Gly Ile Asn Ser Asn Trp Leu
265 270 275 280
30 CCA GTG AAC AAC GCC AAG GTA CCG GAT CCT CGA CCC GGT TCC TGT CAC
Pro Val Asn Asn Ala Lys Val Pro Asp Pro Arg Pro Gly Ser Cys His
285 290 295 300
AAC GAT TCG ACA CCG CTT CCG GAT CCC ACA CTG AAC TTC ATC AAA ACA
Asn Asp Ser Arg Ala Leu Pro Asp Pro Thr Leu Asn Phe Ile Lys Thr
305 310 315 320
35 CAT TCG CTA ATG GAC GAG AAT GTG CCG GCA TTT TTC AGT CAA CCG ATT
His Ser Leu Met Asp Glu Asn Val Pro Ala Phe Phe Ser Gln Pro Ile
325 330 335 340
40 TTG GTC CCG ACG ACC ACA ATA TAC CCG TTC ACT CAA ATC GCC GTA GAT
Leu Val Arg Thr Ser Thr Ile Tyr Arg Phe Thr Gln Ile Ala Val Asp
345 350 355 360
45 GCG CAG ATT AAA ACT CCT GGC GGC AAG ACA TAT GAT GTT ATC TTT GTG
Ala Gln Ile Lys Thr Pro Gly Gly Lys Thr Tyr Asp Val Ile Phe Val
365 370 375 380
50 GGC ACA GAT CAT GGA AAG ATT ATT AAG TCA CTG AAT CCT GAA TCT CCC
Gly Thr Asp His Gly Lys Ile Ile Lys Ser Val Asn Ala Glu Ser Ala
385 390 395 400
GAT TCA CCG GAT AAA CTC ACC TCC GTA GTC ATC GAG CAG ATC GAT GTC
Asp Ser Ala Asp Lys Val Thr Ser Val Val Ile Glu Glu Ile Asp Val
405 410 415 420
55 CTG ACC AAG AGT GAA CCC ATA CCG AAT CTG CAG ATA CTC AGA ACC ATG
Leu Thr Lys Ser Glu Pro Ile Arg Asn Leu Glu Ile Val Arg Thr Met
425 430 435 440
60 CAG TAC GAT CAA CCC AAA GAT GGC ACC TAC CAC GAT GGT AAA TTA ATC
Gln Tyr Asp Gln Pro Lys Asp Gly Ser Tyr Asp Asp Gly Lys Leu Ile
445 450 455 460
65 ATT GTG ACC GAC AGT CAG GTG GTA CCC ATA CAA TTG CAT GGT TGT CAC
Ile Val Thr Asp Ser Gln Val Val Ala Ile Gln Leu His Arg Cys His
465 470 475 480
AAT GAC AAA ATC ACC AGC TGC AGC GAG TCC GCA TTG CAG GAT CCG 1248

WO 9507706

PCT/US94/010151

5 An Asp Lys Ile Thr Ser Cys Ser Glu Cys Val Ala Leu Gln Asp Pro 415
 TAC TGC GCG TGC GAG AAA ATC GGT GGC AGG TGC CGT TCC CAC GCG GGT 1296
 5 Tyr Cys Ala Trp Asp Lys Ile Ala Gly Lys Cys Arg Ser His Gly Ala 430
 420
 CCG CGA TGG CTA GAG GAG AAC TAT TTC TAC CAG AAT GTG GCG ACT GCG 1344
 435
 10 Pro Arg Trp Leu Glu Glu Asp Tyr Phe Tyr Gln Asp Val Ala Thr Gly 445
 CAG CAT GCG GCG TGC CCC TCA GGC AAA ATC AAT TCA AAG GAT GCC AAC 1392
 450
 15 GCT GCG GAG CAG AAG GCG TTC CCG AAC GAC ATC GAC TTA TTG GAT TCG 1440
 465
 20 CCA GCG CAG AGC AAG GAT CAG GAA ATA ATC GAC AAT ATT GAT AAG AAC 1488
 485
 25 TTT GAA GAT ATA ATC AAC GCG CAG TAC ACT GTG GAG ACC CTC GTG ATG 1536
 500
 30 GCC GTT CTG GCG GGT TCG ATC TTT TCG CTG GTC GCG TTC TTT ACA 1584
 515
 35 CCG TAT CCG GAT ACG GAG TAC GAG TAC TTC CAG CAG CEA CAG AAT GTC 1680
 530
 40 AAT AGC TTC CCG TGC TGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT 1728
 545
 45 CCC CAA CTG GAG GAG GGT ACG TAT GCG GAC GCA GTC CTC CTC CCA CAG 1776
 580
 50 CCT CGG CCG CCG AAT AAG ATG CAC TCG CCG AAG AAC ACG CTC GGT AAG 1824
 595
 55 CCG CCG ATG CAG CAG ATG CAG CAG GGT CCC AAC TCG GAG ACC CTC TTC 1872
 610
 60 CAG TTC CAG CTG ACG GGT ACA ACA CCC AGC AGT GGT ATC GTG GTC GCG 1920
 625
 65 ACA ACT TCG CAA CAG TCG GGT CCC ACC ACG TCGGGCGA CAATTACAGG 1970
 645
 CCGCGCGATC GCTTTTCCAC CACCGCGAGC CTCACGAGG TTTACCTTTG AGCGGGAGT 2030
 GGGCGCGTGC AACACGCTCA GCGACTAATT ACCCAATAA TCGCTGTAAA CACACAAAC 2090
 AACCTAAC GAACTTTTGG TCGCGCAGCA AGACAGCGC CCGCTCATCG CATTGTAACT 2150

76

3736

WO 9507706

PCT/US94/010151

CACACCGCT CGAATAGCCG CAGCAGCAG CAGCAGCAGT CCGCAGCAGC CCACTCCAGT 2210
 TCGGCTCCT CCGCGTAAAT GTCAACAGC AGCAGCAGTC CCGCTCCGCG CTCACAGCAGT 2270
 5 CCGAGTCCGC AGCAGAGCCG CAAAGAAATGC AGCTACATCT ACCGTGATTO ATTGATATGC 2330
 AACACAAAT CGATGCCACT CATCCAGGCG CAGTCCAGCG AGCGCCAGCG ACATCCACAC 2390
 10 CCGCAGCGCG ACCCGCTTCC CCGACCCGCT CCGACACCGC CCGCAGCACA CCGACCGCGC 2450
 AGAATGCCAA TCATCGCGAG CACATAGCG AAGTCCATGC CCGTGACACC AGTTCAACGC 2510
 CAATCGCGC TCGCTGAGAC CCGCTCCTAT GAGCTCTAGC AACGCCATCT CGATCGCGCG 2570
 15 ACCTTCCACT TTGGGATCA CGACATATGC GATGATATGC AGCAGACCA CCGAGCAGCG 2630
 TCATCGCTGC CCGATGATCAG ACCCGCGCG CCGTAGGACA CTCGCGATCT GATTGCTGC 2690
 20 CCGCGCTGC CCGCGCTTGC TACATTTGCG TTTCGCAACA CCGAGCTGTT CAGATGAGT 2750
 CCGCGCGAG GTGGAACAC CCGCAGCGCG TCGCAGCGCG AACCGCGCAG CAGCGCCTATC 2810
 AGCCCAAA AGTTGAGTGC GCGCGCAGCG GCGATTTTC CCGCAGCCCA AATCGCCACC 2870
 25 CCACTCAACC GGAAGTGGCG TCATTTGCCA AGAAGCGCG CCGAGCGCGA CAGCAGCTCC 2930
 GCGATTTCTA AGGAGCTCGA CAACTGCTGC CTGCTATGCG TCGACTGCGA TGAGAAATGC 2990
 30 ATGTAATAGA ACCCAACCA CAAATGAGAT AGCAGAACCA CTTCGATTCG CAATTTATAC 3050
 ACCTTTGAT ATTTGAAATA TGACTTCAAT TTAAAAATGC GTAAATATGT TCTATTTT 3110
 TAAGAACCG TTGAGAGAG TTTTCTGCTA CTTAAATAG TACACAAAC TCATATCTAA 3170
 35 CCGCGCTGC CCGATATAGCA ATAAACATCT CCGCTTCCCT TAAATTTAA GTAGCAGTC 3230
 AAAGATCAT TCATATAGCA CAGAAATCTGC ATGCGGATTT ACTTACACAC AAAAGCCAG 3290
 40 AGAATTTATA CAGCAGTTT ATAGTATAT AGCTTTTATA CATACTCCCG GATCTGCTAA 3350
 GTATACACAA CCGAGCATAA CATACATATG CTCTATATAT ACCATATAT 3410
 TTCATAGAG ATTCAATCG ATCGCTAGC CTAAATAGA CCGCAAAAT GATATTTTAA 3470
 45 ATTACGATA CAGAAAAAA AAAGGATTT CCGATATCAG CXTATCGATA CCGTCCAGCT 3530
 CCGNNNGCG CCGCGGTACC CAATTCGCGC 3560

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 650 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:160:

Glu Asp Asp Cys Gln Asp Tyr Ile Arg Ile Met Val Val Pro Ser Pro 1 10 15

Gly Arg Leu Phe Val Cys Gly Thr Asn Ser Phe Arg Pro Met Cys Asn 20 25 30

Thr Tyr Ile Ile Ser Asp Ser Asn Tyr Thr Leu Glu Ala Thr Lys Asn 35 40 45

77

3737

Gly Cln Ala Val Cys Pro Tyr Asp Pro Arg His Asn Ser Thr Ser Val
50 55 60
5 Leu Ala Asp Asn Glu Leu Tyr Ser Gly Thr Val Ala Asp Phe Ser Gly
65 70 75 80
Ser Asp Pro Ile Ile Tyr Arg Cln Pro Leu Cln Thr Glu Cln Tyr Asp
85 90 95
10 Ser Leu Ser Leu Asn Ala Pro Asn Phe Val Ser Ser Phe Thr Cln Gly
100 105 110
Asp Phe Val Tyr Phe Phe Arg Cln Thr Ala Val Cln Phe Ile Asn
115 120 125
15 Cys Gly Lys Ala Ile Tyr Ser Arg Val Ala Arg Val Cys Lys Tyr Asp
130 135 140
Lys Gly Gly Pro His Arg Phe Arg Asn Arg Trp Thr Ser Phe Leu Lys
145 150 155 160
20 Ser Arg Leu Asn Cys Ser Ile Pro Gly Asp Tyr Pro Phe Tyr Phe Asn
165 170 175
25 Glu Ile Cln Ser Ala Ser Asn Leu Val Cln Gly Cln Tyr Gly Ser Met
180 185 190
Ser Ser Lys Leu Ile Tyr Gly Val Phe Asn Thr Pro Ser Asn Ser Ile
195 200 205
30 Pro Gly Ser Ala Val Cys Ala Phe Ala Leu Cln Asp Ile Ala Asp Thr
210 215 220
Phe Glu Gly Cln Phe Lys Glu Cln Thr Gly Ile Asn Ser Asn Trp Leu
225 230 235 240
Pro Val Asn Asn Ala Lys Val Pro Asp Pro Arg Pro Gly Ser Cys His
245 250 255
40 Asn Asp Ser Arg Ala Leu Pro Asp Pro Thr Leu Asn Phe Ile Lys Thr
260 265 270
His Ser Leu Met Asp Glu Asn Val Pro Ala Phe Phe Ser Cln Pro Ile
275 280 285
45 Leu Val Arg Thr Ser Thr Ile Tyr Arg Phe Thr Cln Ile Ala Val Asp
290 295 300
Ala Cln Ile Lys Thr Pro Gly Gly Lys Thr Tyr Asp Val Ile Phe Val
305 310 315 320
Gly Thr Asp His Gly Lys Ile Ile Lys Ser Val Asn Ala Glu Ser Ala
325 330 335
55 Asp Ser Ala Asp Lys Val Thr Ser Val Val Ile Glu Glu Ile Asp Val
340 345 350
Leu Thr Lys Ser Glu Pro Ile Arg Asn Leu Glu Ile Val Arg Thr Met
355 360 365
60 Cln Tyr Asp Cln Pro Lys Asp Gly Ser Tyr Asp Asp Gly Lys Leu Ile
370 375 380
Ile Val Thr Asp Ser Cln Val Val Ala Ile Cln Leu His Arg Cys His
385 390 395 400
Asn Asp Lys Ile Thr Ser Cys Ser Glu Cys Val Ala Leu Cln Asp Pro
405 410 415

78

3738

Tyr Cys Ala Trp Asp Lys Ile Ala Gly Lys Cys Arg Ser His Gly Ala
420 425 430
5 Pro Arg Trp Leu Glu Cln Asn Tyr Phe Tyr Cln Asn Val Ala Thr Gly
435 440 445
Cln His Ala Ala Cys Pro Ser Gly Lys Ile Asn Ser Lys Asp Ala Asn
450 455 460
10 Ala Gly Glu Cln Lys Gly Phe Arg Asn Asp Met Asp Leu Leu Asp Ser
465 470 475 480
Arg Arg Cln Ser Lys Asp Cln Glu Ile Ile Asp Asn Ile Asp Lys Asn
485 490 495
15 Phe Glu Asp Ile Ile Asn Ala Cln Tyr Thr Val Glu Thr Leu Val Met
500 505 510
Ala Val Leu Ala Gly Ser Ile Phe Ser Leu Leu Val Gly Phe Thr
515 520 525
20 Gly Tyr Phe Cys Gly Arg Arg Cys His Lys Asp Glu Asp Asn Leu
530 535 540
25 Pro Tyr Pro Asp Thr Glu Tyr Glu Tyr Phe Glu Cln Arg Cln Asn Val
545 550 555 560
Asn Ser Phe Pro Ser Ser Cys Arg Ile Cln Cln Glu Pro Lys Leu Leu
565 570 575
30 Pro Cln Val Glu Glu Val Thr Tyr Ala Asp Ala Val Leu Leu Pro Cln
580 585 590
Pro Pro Pro Pro Asn Lys Met His Ser Pro Lys Asn Thr Arg Lys
595 600 605
35 Pro Pro Met His Cln Met His Cln Cln Gly Pro Asn Ser Glu Thr Leu Phe
610 615 620
Cln Phe His Val Thr Ala Thr Thr Pro Ser Ser Arg Ile Val Val Ala
625 630 635 640
Thr Thr Ser Glu His Cys Val Pro Thr Arg
645 650

(2) INFORMATION FOR SEQ ID NO161:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2670 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(1x) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 268..2439

(1x) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GAATATCGAA CCCCAGATTG AATGACACGC AAAACGCCAA TTAGATAGTT GCAAGCTTAA 60
TGCATTTTCAG AATTTTTHHC GATGCGAACC AAGTTCCGCC ACGAAGTCA ACGTGTGTA 120
AATGCCACAC AATCTCGACG CCAACACCA AACACAAAG AACACGAC CCGCTTCTAC 180

79

3739

TCCGCTCTCC ACTTATATCC AATTGAGGTT GGTGGGGTGG CATTCGGCCCG CCGGCTGACC 240
 ACCGCTCTCG CTCGACCGCG CCGTGGCA ATG TCT CTA CAG CTA TCG CCG CTC 294
 Met Ser Leu Leu Gln Leu Ser Pro Leu
 5 CTC GCA CTC CTA CTC TCG AGT AGT GTG AGC GAG AGC GGT CCG 342
 Leu Ala Leu Leu Leu Cys Ser Ser Val Ser Gln Thr Ala Ala 25
 10 GAC TAC GAG AAC ACC TGG AAC TTC TAC TAC GAG CCG CCG TGT TGC ACT 390
 Asp Tyr Glu Asn Thr Trp Asn Phe Tyr Tyr Glu Arg Pro Cys Cys Thr 40
 15 GGA AAC GAT CAG GCG AAC AAC AAT TAC GGA AAA CAC GCG CCA GAT CAT 438
 Gly Asn Asp Gln Gly Asn Asn Asn Tyr Gly Lys His Gly Ala Asp His 55
 CTC GCG GAG TTC AAC TGC GCG AGC TAC TAC TAT CCG ACA TTC CAT ATG 486
 Val Arg Glu Phe Asn Cys Gly Lys Leu Tyr Tyr Arg Thr Phe His Met 70
 20 AAC GAA GAT CCA GAT ACG CTC TAT TAT GTG GGA CCG ATG GAT CCG GTA TTC 534
 Asn Glu Asp Arg Asp Thr Leu Tyr Val Gly Ala Met Asp Arg Val Phe 85
 CGT GTG AAC CTG CAG AAT ATC TCC TCA TCC AAT TGT AAT CCG GAT CCG 582
 Arg Val Asn Leu Gln Asn Ile Ser Ser Ser Asn Cys Asn Arg Asp Ala 105
 30 ATC AAC TTG GAG CCA ACA CCG GAT GAT GTG GGT AGC TGC GTC TCC AAA 630
 Ile Asn Leu Glu Pro Thr Arg Asp Asp Val Val Ser Cys Val Ser Lys 120
 GGC AAA AGT CAG ATC TTC GAC TGC AAC AAC CAT GTG CCG ATC ATC CAG 678
 Gly Lys Ser Gln Ile Phe Asp Cys Lys Asn His Val Arg Val Ile Gln 135
 TCA ATG CAC GAG GCG GAT AGC CTC TAT GTA TGC GGC ACC AAC GCC CAC 726
 Ser Met Asp Gln Gly Asp Arg Leu Tyr Val Cys Gly Thr Asn Ala His 150
 40 AAT CCC AAG GAT TAT GTT ATC TAT TGC AAT CTA ACC CAC CTG CCG CCG 774
 Asn Pro Lys Asp Tyr Val Ile Tyr Ala Asn Leu Thr His Leu Pro Arg 165
 TCG GAA TAT GTG ATT GCG GTG GGT CTG GCG ATT GCC AAG TGC CCC TAC 822
 Ser Glu Tyr Val Ile Gly Val Gly Leu Gly Ile Ala Lys Cys Pro Tyr 185
 50 GAT CCC CTC CAC AAC TCA ACT CCG ATT TAT GTG GAG AAT GCG AAT CCG 870
 Asp Pro Leu Asp Asn Ser Thr Ala Ile Tyr Val Glu Asn Gly 200
 GGT GGT CTG CCG GGT TTG TAC TCC GCG ACC AAT CCG GAG TTG ACC AAG 918
 Gly Gly Leu Pro Gly Leu Tyr Ser Tyr Glu Thr Asn Ala Glu Phe Thr Lys 215
 60 CCG GAT ACG GTT ATT TTC CCG ACT GAT CTG TAT AAT ACT TCG CCG AAA 966
 Ala Asp Thr Val Ile Phe Arg Thr Asp Leu Tyr Asn Thr Ser Ala Lys 230
 CGT TTG GAA TAT AAA TTC AAG AGG ACT CTG AAA TAC CAC TCC AAG TCG 1014
 Arg Leu Glu Tyr Lys Phe Lys Arg Thr Leu Lys Tyr Asp Ser Lys Trp 245
 65 TTG GAC AAA CCA AAC TTT CTC GCG TCC TTT GAT ATT GCG GAG TAC GTG 1062

Leu Asp Lys Pro Asn Phe Val Gly Ser Phe Asp Ile Gly Glu Tyr Val 250
 5 TAT TTC TTT TTC COT GAA ACC CCG CTG GAA TAC ATC AAC TGC CCG AAG 1110
 Tyr Phe Phe Phe Arg Glu Thr Ala Val 275
 GCT GTC TAT TCG CCG ATC GCA CCG GTG TGC AAG AAG GAT GTG CGT GCA 1158
 Ala Val Tyr Ser Arg Ile Ala Arg 290
 10 AAG AAT CTG CTG CCG CAC AAC TGG CCG ACC TAC CTG AAG GCG AAG CTC 1206
 Lys Asn Leu Leu Ala His Asn Trp Ala Thr Tyr Leu Lys Ala Arg Leu 310
 15 AAC TGC AGC ATC TCC GCG GAA TTT CCG TTC TAT TTC AAC GAG ATC CAA 1354
 Asn Cys Ser Ile Ser Gly Glu Phe Pro Phe Tyr Phe Asn Glu Ile Gln 325
 20 TCG GTC TAC CAG CTG CCG TCC GAT AAG AGT CGA TTC TTC GCC ACA TTC 1302
 Ser Val Tyr Gln Leu Pro Ser Asp Lys Ser Arg Phe Ala Thr Phe 340
 25 ACG ACG ACG ACT AAT GCG CTG ATT GCG TCT GCG GTA TGC AGT TTC CAC 1350
 Thr Thr Ser Thr Asn Gly Leu Ile Gly Ser Ala Val Cys Ser Phe His 350
 30 AAT AAC GAG ATT CAG GCT CCG TTC AAT GCG AAA TTC AAG CAG CAA TCT 1398
 Ile Asn Glu Ile Gln Ala Phe Asn Gly Lys Phe Lys Glu Gln Ser 375
 TCA TCG AAT TCC GCA TCG CTG CCG GTG CTT AAC TCC CCG GTG CCG GAA 1446
 Ser Ser Asn Ser Ala Trp Leu Pro Val Leu Asn Ser Arg Val Pro Glu 390
 35 CCA CCG CCG GGT ACA TGT GTC AAC GAT ACA TCA AAC CTG CCG GAT ACC 1494
 Pro Arg Pro Gly Thr Cys Val Asn Asp Thr Ser Asn Leu Pro Asp Thr 405
 40 GTA CTG AAT TTC ATC AGA TCC CAT CCA CTT ATG GAC AAA GCC GTA AAT 1542
 Val Leu Asn Phe Ile Arg Ser His Pro Leu Met Asp Lys Ala Val Asn 425
 45 CAC GAG CAC AAC AAT CCA GTC TAT TAT AAA AGG GAT TTG GTC TTC ACC 1590
 His Glu His Asn Asn Pro Val Tyr Tyr Lys Arg Asp Leu Val Phe Thr 435
 AAG CTC GTC GTT CAC AAA ATT CCG ATT CCG ATC CTC AAC CAG GAA TAC 1638
 Lys Leu Val Val Asp Lys Ile Arg Ile Asp Ile Leu Asn Gln Glu Tyr 455
 50 ATT CTG TAC TAT GTG GCG ACC AAT CTG GGT CCG ATT TAC AAA ATC GTG 1686
 Ile Val Tyr Tyr Val Gly Thr Asn Leu Gly Arg Ile Tyr Lys Ile Val 470
 55 CAG-TAC TAC COT AAC GGA GAG TCG CTG TCC AAG CTT CTG GAT ATC TTC 1734
 Gln Tyr Tyr Arg Asn Gly Glu Ser Leu Ser Lys Leu Leu Asp Ile Phe 485
 60 GAG GTG GCT CCA AAC GAG CCG ATC CAA GTG ATG GAA ATC AGC CAG ACA 1782
 Glu Val Ala Pro Asn Glu Ala Ile Gln Val Met Glu Ile Ser Gln Thr 490
 65 CGT AAG ACG CTC TAC ATT GCG ACG GAT CAT CCG ATC AAG CAA ATC CAC 1830
 Arg Lys Ser Leu Tyr Ile Gly Thr Asp His Arg Ile Lys Gln Ile Asp 515
 CTG CCG ATG TGC AAT CCG COT TAC CAC AAC TGC TTC CCG TCC GGT COT 1878

Leu Ala Met Cys Asn Arg Arg Tyr Asp Asn Cys Phe Arg Cys Val Arg		525	
GAT CCC TAC TCG CGC GAT AAG GAG GCC AAT ACG TGC CGA CCG TAC		1926	
5	Asp Pro Tyr Cys Gly Trp Asp Lys Glu Ala Asn Thr Cys Arg Pro Tyr	540	
GAG CTG GAT TTA CTG CAG GAT GTG GCC AAT GAA ACG AGT GAC ATT TGC		1974	
10	Glu Leu Asp Leu Leu Gln Asn Val Ala Asn Glu Thr Ser Asp Ile Cys	555	
GAT TCG ACT CTG CTG AAA AAG AAG ATT CTG ACC TAT GCC CAG AGT		2028	
Asp Ser Ser Val Leu Leu Lys Lys Ile Val Val Thr Tyr Gly Gln Ser		570	
15	GTA CAT CTG GGC TGT TTC GTC AAA ATA CCC GAA CTG CAG AAT GAG	2070	
Val His Leu Gly Cys Phe Val Lys Ile Pro Glu Val Leu Lys Asn Glu		590	
20	CAA CTG ACC TCG TAT CAT CAC TCC AAG GAC AAG GGA CCG TAC GAG ATT	2118	
Gln Val Thr Tyr Tyr His His Ser Lys Asp Lys Gly Arg Tyr Glu Ile		605	
25	CCT TAC TCG CCG ACC AAA TAC ATT GAG ACC ACC GAA CCG CTG CTT	2166	
Arg Tyr Ser Pro Thr Lys Tyr Ile Glu Thr Thr Glu Arg Gly Leu Val		620	
30	GTG GTT TCG GTC AAC GAA GCC GAT GGT GGT CCG TAC GAT TGC CAT TTG	2214	
Val Val Ser Val Asn Glu Ala Asp Gly Gly Arg Tyr Asp Cys His Leu		635	
35	CGC CGC TCG CTT TTG TCG ACC TAC AAC ATT ACA CTG GAT GCC CAC AGA	2262	
Gly Cys Ser Leu Leu Cys Ser Tyr Asn Ile Thr Val Asp Ala His Arg		650	
40	TGC ACT CCG CCG AAC AAG AGT AAT GAC TAT CAG AAA ATC TAC TCG GAC	2310	
Cys Thr Pro Pro Asn Lys Ser Asn Asp Tyr Glu Lys Ile Tyr Ser Asp		670	
45	TGC TGC CAC GAG TTC GAG AAA TAC AAA ACA GCA ATG AAG TCC TCG GAA	2358	
Trp Cys His Glu Phe Glu Lys Tyr Lys Thr Ala Met Lys Ser Trp Glu		685	
50	AAG AAG CAA GGC CAA TCG TCG ACA CCG CAG AAC TTC ACG TCC AAT CAG	2406	
Lys Lys Gln Gly Gln Cys Ser Thr Arg Gln Asn Phe Ser Cys Asn Gln		700	
55	CAT CCG AAT GAG ATT TTC CCG AAG CCC AAT GTC TCATATCAGC AAGACAGTAT	2459	
His Pro Asn Glu Ile Phe Arg Lys Pro Asn Val		715	
60	CGCCCTCMAA ATGCCGTGAT CCGCTCCAA TCATTTTAG TTAATCGAA GCGAGAGGA	2519	
TAAATACAGT GCGGATAGA AACCCAGGA CCGAGAGAC TCATTTATAT CATTTATATC		2579	
ACCGACATCA TCATAGACAT ACTTTCTCA GCAATCAACA GAAACTCTT CTTAAGCAT		2636	
TATGCATTTA CCGAAGCATT TACATGCAT C		2670	

(2) INFORMATION FOR SEQ ID NO:62:

(1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 724 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Met Ser Leu Leu Gln Leu Ser Pro Leu Leu Ala Leu Leu Leu Leu		1	
5	Cys Ser Ser Val Ser Glu Thr Ala Ala Asp Tyr Glu Asn Thr Trp Asn	20	
10	Phe Tyr Tyr Glu Arg Pro Cys Cys Thr Gly Asn Asp Gln Gly Asn Asn	35	
Asn Tyr Gly Lys His Gly Ala Asp His Val Arg Glu Phe Asn Cys Gly		50	
15	Lys Leu Tyr Tyr Arg Thr Phe His Met Asn Glu Asp Arg Asp Thr Leu	65	
Tyr Val Gly Ala Met Asp Arg Val Phe Arg Val Asn Leu Gln Asn Ile		80	
20	Ser Ser Ser Asn Cys Asn Arg Asp Ala Ile Asn Leu Glu Pro Thr Arg	100	
25	Asp Asp Val Val Ser Cys Val Ser Lys Gly Lys Ser Gln Ile Phe Asp	115	
Cys Lys Asn His Val Arg Val Ile Gln Ser Met Asp Gln Gly Asp Arg		130	
30	Leu Tyr Val Cys Gly Thr Asn Ala His Asn Pro Lys Asp Tyr Val Ile	145	
35	Tyr Ala Asn Leu Thr His Leu Pro Arg Ser Glu Tyr Val Ile Gly Val	160	
Gly Leu Gly Ile Ala Lys Cys Pro Tyr Asp Pro Leu Asp Asn Ser Thr		175	
40	Ala Ile Tyr Val Glu Asn Gly Asn Pro Gly Leu Pro Gly Leu Tyr	190	
Ser Gly Thr Asn Ala Glu Phe Thr Lys Ala Asp Thr Val Ile Phe Arg		205	
45	Thr Asp Leu Tyr Asn Thr Ser Ala Lys Arg Leu Glu Tyr Lys Phe Lys	220	
Arg Thr Leu Lys Tyr Asp Ser Lys Trp Leu Asp Lys Pro Asn Phe Val		235	
50	Gly Ser Phe Asp Ile Gly Glu Tyr Val Tyr Phe Phe Arg Glu Thr	250	
55	Ala Val Glu Tyr Ile Asn Cys Gly Lys Ala Val Tyr Ser Arg Ile Ala	265	
Arg Val Cys Lys Lys Asp Val Gly Gly Lys Asn Leu Ala His Asn		280	
60	Trp Ala Thr Tyr Leu Lys Ala Arg Leu Asn Cys Ser Ile Ser Gly Glu	295	
Phe Pro Phe Tyr Phe Asn Glu Ile Gln Ser Val Tyr Gln Leu Pro Ser		310	
65	Asp Lys Ser Arg Phe Phe Ala Thr Phe Thr Thr Ser Thr Asn Gly Leu	325	

WO 95/07706

Lys Pro Asn Val

(2) INFORMATION FOR SEQ ID NO:63:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2504 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(12) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: 355..2493

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:63:

60 GCGCGGTGGA CACAGAGGGA AGTTAGTAT CAACTGAGA GTTCTTTGG ACCGATTTT
 120 ACGGAGGGA CATTAAAT TCGGACAAA TCGTCTTTG GTGCTTCTT GTGATTTT
 180 GTGTTCTGA AGATGCTCC CTGCTTTTC GGAATAGCTT TCGTGTGAT TGTGTGTT
 240 TTGAGATGC TTGCTTCTT TTTCGATTA GCTTTCGAG GTGCTTTGAG CCGCGCTTG
 300 TTGACCCGC CACATAATCT TCGACTACA ATGAGAGGA AATTGAAA CCGCTTCAG
 357 ACCGTACA TCGACAAAT GTTGTCTTC CAATGATCT TCGAATGAG CTAC ATC
 Met
 1
 405 GTG GTG AAG ATC TTG GTT TCG TCG ATA TGT CTG ATA GCG CTG TGT CAT
 15 Val Val Lys Ile Leu Val Trp Ser Ile Cys Leu Ile Ala Leu Cys His
 453 GCT TCG ATG CCG GAT AGT TCT TCG AAA TTA ATA AAC CAT TTT AAA TCA
 30 Ala Trp Met Pro Asp Ser Ser Lys Leu Ile Asn His Phe Lys Ser
 501 GTT GAA ACT AAA AGC TTT ACC GCG AAC GCG AGC TTC CCT CAT CAC TTT
 45 Val Glu Ser Lys Ser Phe Thr Gly Asn Ala Thr Phe Pro Asp His Phe
 549 ATT GTC TTG AAT CAA CAC GAA ACT TCG ATA TTA GTC GGT AGA AAT
 60 Ile Val Leu Asn Gln Asp Glu Thr Ser Ile Leu Val Gly Gly Arg Asn
 597 AGG GTT TAC AAT TTA ACT ATA TTC GAC CTC AGT GAG CCG AAT AAA GCG GCG
 75 Arg Val Tyr Asn Leu Ser Ile Phe Asp Leu Ser Glu Arg Lys Gly Gly
 645 CGA ATC GAC TCG CCA TCG GAT GCA CAT GCG CAG TTG TGT ATA TTC
 90 Arg Ile Asp Trp Pro Ser Ser Asp Ala His Gly Gln Leu Cys Ile Leu
 693 AAA GCG AAA ACG GAC GAC TCG CAA AAT TAC ATT AGA ATA CTG TAC
 105 Lys Gly Lys Thr Asp Asp Asp Cys Gln Asn Tyr Ile Arg Ile Leu Tyr
 741 TCT TCA CAA CCG GCG AAA TTA GTT ATT TCG GCG ACC AAT TCG TAC AAA
 120 Ser Ser Glu Pro Gly Lys Leu Val Ile Cys Gly Thr Asn Ser Tyr Lys
 789 CCC CTC TGT CCG ACG TAC GCA TTT AAG CAG GCA AAG TAC CTG GTT CAG
 135 Pro Leu Cys Arg Thr Tyr Ala Phe Lys Glu Gly Lys Tyr Leu Val
 140 130 145

85

3745

PCT/US94/10151

WO 95/07706

Ile Gly Ser Ala Val Cys Ser Phe His Ile Asn Glu Ile Gln Ala Ala
 365 360
 5 Phe Asn Gly Lys Phe Lys Glu Gln Ser Ser Asn Ser Ala Trp Leu
 370 375
 Pro Val Leu Asn Ser Arg Val Pro Glu Pro Arg Pro Gly Thr Cys Val
 385 390 395
 10 Asn Asp Thr Ser Asn Leu Pro Asp Thr Val Leu Asn Phe Ile Arg Ser
 405 410 415
 His Pro Leu Met Asp Lys Ala Val Asn His Glu His Asn Asn Pro Val
 420 425
 15 Tyr Tyr Lys Arg Asp Leu Val Phe Thr Lys Leu Val Val Asp Lys Ile
 435 440 445
 20 Arg Ile Asp Ile Leu Asn Gln Glu Tyr Ile Val Tyr Tyr Val Gly Thr
 450 455 460
 Asn Leu Gly Arg Ile Tyr Lys Ile Val Gln Tyr Tyr Arg Asn Gly Glu
 465 470 475
 25 Ser Leu Ser Lys Leu Leu Asp Ile Phe Glu Val Ala Pro Asn Glu Ala
 485 490 495
 Ile Gln Val Met Glu Ile Ser Gln Thr Arg Lys Ser Leu Tyr Ile Gly
 500 505 510
 30 Thr Asp His Arg Ile Lys Gln Ile Asp Leu Ala Met Cys Asn Arg Arg
 515 520 525
 Tyr Asp Asn Cys Phe Arg Cys Val Arg Asp Pro Tyr Cys Gly Trp Asp
 530 535 540
 35 Lys Glu Ala Asn Thr Cys Arg Pro Tyr Glu Leu Asp Leu Leu Gln Asp
 545 550 555
 40 Val Ala Asn Glu Thr Ser Asp Ile Cys Asp Ser Ser Val Leu Lys Lys
 565 570 575
 Lys Ile Val Val Thr Tyr Gly Gln Ser Val His Leu Gly Cys Phe Val
 580 585 590
 45 Lys Ile Pro Glu Val Leu Lys Asn Glu Gln Val Thr Trp Tyr His His
 595 600 605
 Ser Lys Asp Lys Gly Arg Tyr Glu Ile Arg Tyr Ser Pro Thr Lys Tyr
 610 615 620
 50 Ile Glu Thr Thr Glu Arg Gly Leu Val Val Ser Val Asn Glu Ala
 625 630 635
 55 Asp Gly Gly Arg Tyr Asp Cys His Leu Gly Gly Ser Leu Leu Cys Ser
 645 650 655
 Tyr Asn Ile Thr Val Asp Ala His Arg Cys Thr Pro Asn Lys Ser
 660 665 670
 60 Asn Asp Tyr Gln Lys Ile Tyr Ser Asp Trp Cys His Glu Phe Glu Lys
 675 680 685
 65 Tyr Lys Thr Ala Met Lys Ser Trp Glu Lys Lys Gln Gly Gln Cys Ser
 690 695 700
 Thr Arg Gln Asn Phe Ser Cys Asn Gln His Pro Asn Glu Ile Phe Arg
 705 710 715 720

84

3744

AAA GAA GTA GAG ATA GCG TTT TGT TCA TAC AAT CCG GAA CAC AAC Lys Glu Val Glu Gly Ile Gly Leu Cys Pro Tyr Aen Pro Glu His Aen 150 155		837
5 AGC ACA TCT GTC TCC TAC AAT GCG CAA TTA TTT TCA GCG ACG GTC GCG Ser Thr Ser Val Ser Tyr Aen Gly Gln Leu Phe Ser Ala Thr Val Ala 165 170		885
10 GAC TTT TCC GCG GCG GAC CCT CTC ATA TAC AGG GAG CCG CAG CCG ACC Aap Phe Ser Gly Gly Aap Pro Leu Ile Tyr Arg Glu Pro Gln Arg Thr 180 185		933
15 GAA CTC TCA GAT CTC AAA CAA CTG AAC GCA CCG AAT TTC GTA AAC TCG Glu Leu Ser Asp Leu Lys Gln Leu Aen Ala Pro Aen Phe Val Aen Ser 195 200		981
20 GAG TAC ATG AAG TGC GGA AAA GTC ATC TAC TCG CCG GTC GCC AGG GTG Glu Tyr Met Aen Cys Gly Tyr Ile Phe Phe Phe Phe Phe Val Ala Val 210 215		1029
25 TGC AAG GAC GAC AAA GCG GCG CCT CAC CAG TCA CCG GCG TCG ACG Cys Lys Aap Asp Lys Gly Gly Pro His Gln Ser Arg Asp Arg Trp Thr 245 250		1125
30 TCG TTC CTC AAA CCA COT CTC AAT TGT TCA ATT CCC GCG GAG TAC CCG Ser Phe Leu Lys Ala Arg Leu Aen Cys Ser Ile Pro Gly Glu Tyr Pro 260 265		1173
35 TTT TAC TTT GAT GAA ATC CAA TCA ACA AGT GAT ATA GTC GAG GGT CCG Phe Tyr Phe Asp Glu Ile Gln Ser Thr Ser Asp Ile Val Glu Gly Arg 275 280		1221
40 TAC AAT TCC GAC GAC AGC AAA AAG ATC ATT TAT GGA ATC CTC ACA ACT Tyr Aen Ser Asp Asp Ser Lys Lys Ile Ile Tyr Gly Ile Leu Thr Thr 290 295		1269
45 GAC ATC TTG CCG GTG TTT GAA GGG AGC TTC AAG CAC CAA GAG ACG ATC Aap Ile Leu Arg Val Phe Glu Gly Ser Phe Lys His Gln Glu Thr Ile 310 315		1365
50 AAC TCG AAC TCG CTC CCG GCG CCG CAC CTA GTC CCT GAA CCC ACG Aen Ser Aen Trp Leu Pro Val Pro Gln Aen Leu Val Pro Glu Pro Arg 340 345		1413
55 CCC GCG CAG TGC GTA CCG GAC AGC AGG ATC CTC CCG GAC AAG AAC GTC Pro Gly Gln Cys Val Arg Asp Ser Arg Ile Leu Pro Asp Lys Aen Val 355 360		1461
60 AAC TTT ATT AAG ACC CAG TCT TTG ATG CAG GAC GTT CCG GCT GTT TTC Aen Phe Ile Lys Thr His Ser Leu Met Glu Asp Val Pro Ala Leu Phe 370 375		1509
65 GGA AAA CCA GTT CTC GTC CCA GTG ACT CTC CAG TAT CCG TTT ACA GCC Gly Lys Pro Val Leu Val Arg Val Ser Leu Gln Tyr Arg Phe Thr Ala 390 395		1557
ATA ACA CTG GAT CCA CAA CTG AAA ACA ATC AAT AAT CAG TAT CTC GAT Ile Thr Val Aap Pro Gln Val Lys Thr Ile Aen Aen Gln Tyr Leu Aap 405 410		1605
GTT TTG TAT ATC GGA ACA GAT CAT CCG AAG GTA CTA AAA GCT GTT AAT Val Leu Tyr Ile Gly Thr Asp Asp Gly Lys Val Leu Lys Ala Val Aen 420 425		1653
5 ATA CCA AAG CCA CAC GCT AAA CCG TTG TTA TAT CGA AAA TAC GGT ACA Ile Pro Lys Arg His Ala Lys Ala Leu Leu Tyr Arg Lys Tyr Arg Thr 435 440		1701
10 TCC GTA CAT CCG CAC GGA GCT CCC GTA AAA CAG CTG AAG ATC GCT CCC Ser Val His Pro His Gly Ala Pro Val Lys Gln Leu Lys Ile Ala Pro 450 455		1749
15 GGT TAT GCG AAA GTT GTG GTC GCG AAA GAC GAA ATC AGA CTT GCT Gly Tyr Gly Lys Val Val Val Val Val Gly Lys Asp Glu Ile Arg Leu Ala 470 475		1797
20 AAT CTC AAC CAT TGT CCA ACC AAA ACG CCG TCC AAG CAC TGT GTG GAA Aen Leu Aen His Cys Ala Ser Lys Thr Arg Cys Lys Asp Cys Val Glu 485 490		1845
CTG CAA GAC CCA CAT TGC CCG TCG GAC GCG AAA CAA AAC CTG TGT GTC Leu Gln Aap Pro His Cys Ala Trp Aap Ala Lys Gln Aen Leu Cys Val 500 505		1893
25 ACC ATT GAC ACC GTC ACT TCG TAT CCG TTC CTG ATC CAG GAC GTA GTT Ser Ile Asp Thr Val Thr Ser Tyr Arg Phe Leu Ile Gln Aap Val Val 515 520		1941
30 CCG GCG GAC GAC AAC AAA TGT TCG TCG CCG CAA ACA GAC AAA AAG ACT Arg Gly Aap Asp Aen Lys Cys Tyr Ser Pro Gln Thr Aap Lys Lys Thr 530 535		1989
35 GTG ATT AAG AAT AAG CCG AGC GAG GTT GAG AAC GAG ATT ACG AAC TCC Val Ile Lys Aen Lys Pro Ser Glu Val Glu Aen Glu Ile Thr Aen Ser 550 555		2037
40 ATT GAC GAA AAG GAT CTC CAT TCA ACC GAT CCG CTC ATC AAA ACT GGT Ile Aap Glu Lys Aap Leu Asp Ser Ser Aap Pro Leu Ile Lys Thr Gly 565 570		2085
CTC GAT CAC GAT TCC GAT TGT GAT CCA GTC ACG CAG AAC ACG ATA GCG Leu Aap Asp Aap Ser Asp Cys Aap Pro Val Ser Glu Aen Ser Ile Gly 580 585		2133
45 GGA TCG CCG GTC CCG CAG CAA CTT GTT ATA TAC ACA GCT GCG ACT CTA Gly Cys Ala Val Arg Gln Gln Leu Val Ile Tyr Thr Ala Gly Thr Leu 595 600		2181
50 CAC ATT CTC GTC GTC GTC AGC ATC GTG GGT TTA TTT TCT TCG CTT His Ile Val Val Val Val Val Val Val Val Val Val Val Val Val Val 610 615		2229
TAT AGC GGG TTA TCT GTT TTC CCA AAA TTT CAC TCG GAT TCG CAA TAT Tyr Ser Gly Leu Ser Val Phe Ala Lys Phe His Ser Aap Ser Gln Tyr 630 635		2277
CCT GAG CCG CCG TTT ATA GAG CAG CAC AAT CAT TTG GAA AGA TTA ACG Pro Glu Ala Pro Phe Ile Glu Gln His Aen His Leu Glu Arg Leu Ser 645 650		2325
GCC AAC CAG ACG GCG TAT TTG ACT CCG AGG CCG AAT AAA CCG GTC AAT Ala Aen Gln Thr Gly Tyr Leu Thr Pro Arg Ala Aen Lys Ala Val Aen 660 665		2373
TTG CTG GTC AAG GTG TCT AGT ACG CCG CCG CCG AAA AAG GAC AAT Leu Val Val Lys Val Ser Ser Thr Pro Arg Pro Lys Lys Aap Aen 675 680		2421

CTC GAT GTC AGC AAA GAC TTG AAC ATT GCG AGT CAC GGG ACT TTT CAA 2469
 Leu Asp Val Ser Lys Asp Leu Asn Ile Ala Ser Asp Gly Thr Leu Gln 705
 690
 5 AAA ATC AAG AAG ACT TAC ATT TAGTCGAGCT TTTT 2504
 Lys Ile Lys Lys Thr Tyr Ile 710

10 (2) INFORMATION FOR SEQ ID NO:164:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 712 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(12) SEQUENCE DESCRIPTION: SEQ ID NO:164:

Met Val Val Lys Ile Leu Val Trp Ser Ile Cys Leu Ile Ala Leu Cys 15
 1
 His Ala Trp Met Pro Asp Ser Ser Ser Lys Leu Ile Asn His Phe Lys 30
 20
 Ser Val Glu Ser Lys Ser Phe Thr Gly Asn Ala Thr Phe Pro Asp His 45
 35
 Phe Ile Val Leu Asn Gln Asp Glu Thr Ser Ile Leu Val Gly Gly Arg 60
 50
 Asn Arg Val Tyr Asn Leu Ser Ile Phe Asp Leu Ser Glu Arg Lys Gly 80
 65
 Gly Arg Ile Asp Trp Pro Ser Ser Asp Ala His Gly Gln Leu Cys Ile 95
 80
 Leu Lys Gly Lys Thr Asp Asp Asp Cys Gln Asn Tyr Ile Arg Ile Leu 110
 100
 Tyr Ser Ser Glu Pro Gly Lys Leu Val Ile Cys Gly Thr Asn Ser Tyr 125
 115
 Lys Pro Leu Cys Arg Thr Tyr Ala Phe Lys Glu Gly Lys Tyr Leu Val 140
 130
 Glu Lys Glu Val Glu Gly Ile Gly Leu Cys Pro Tyr Asn Pro Glu His 160
 145
 Asn Ser Thr Ser Val Ser Tyr Asn Gly Gln Leu Phe Ser Ala Thr Val 175
 165
 Ala Asp Phe Ser Gly Gly Asp Pro Leu Ile Tyr Arg Glu Pro Gln Arg 190
 180
 Thr Glu Leu Ser Asp Leu Lys Gln Leu Asn Ala Pro Asn Phe Val Asn 205
 195
 Ser Val Ala Tyr Gly Asp Tyr Ile Phe Phe Thr Arg Glu Thr Ala 220
 210
 Val Glu Tyr Met Asn Cys Gly Lys Val Ile Tyr Ser Arg Val Ala Arg 240
 225
 Val Cys Lys Asp Asp Lys Gly Gly Pro His Gln Ser Arg Asp Arg Trp 255
 245

Thr Ser Phe Leu Lys Ala Arg Leu Asn Cys Ser Ile Pro Gly Glu Tyr 260
 265
 Pro Phe Tyr Phe Asp Glu Ile Gln Ser Thr Ser Asp Ile Val Glu Gly 285
 275
 Arg Tyr Asn Ser Asp Asp Ser Lys Lys Ile Ile Tyr Gly Ile Leu Thr 300
 290
 10 Thr Pro Val Asn Ala Ile Gly Cys Ser Ala Ile Cys Ala Tyr Gln Met 315
 305
 Ala Asp Ile Leu Arg Val Phe Glu Gly Ser Phe Lys His Gln Glu Thr 335
 325
 Ile Asn Ser Asn Trp Leu Pro Val Pro Gln Asn Leu Val Pro Glu Pro 350
 340
 20 Arg Pro Gly Gln Cys Val Arg Asp Ser Arg Ile Leu Pro Asp Lys Asn 365
 355
 Val Asn Phe Ile Lys Thr His Ser Leu Met Glu Asp Val Pro Ala Leu 380
 370
 Phe Gly Lys Pro Val Leu Val Arg Val Ser Leu Gln Tyr Arg Phe Thr 400
 385
 Ala Ile Thr Val Asp Pro Gln Val Lys Thr Ile Asn Asn Gln Tyr Leu 415
 405
 Asp Val Leu Tyr Ile Gly Thr Asp Asp Gly Lys Val Leu Lys Ala Val 430
 420
 35 Asn Ile Pro Lys Arg His Ala Lys Ala Leu Leu Tyr Arg Lys Tyr Arg 445
 435
 Thr Ser Val His Pro His Gly Ala Pro Val Lys Gln Leu Lys Ile Ala 460
 450
 40 Pro Gly Tyr Gly Lys Val Val Val Val Gly Lys Asp Glu Ile Arg Leu 480
 465
 Ala Asn Leu Asn His Cys Ala Ser Lys Thr Arg Cys Lys Asp Cys Val 495
 485
 Glu Leu Gln Asp Pro His Cys Ala Trp Asp Ala Lys Gln Asn Leu Cys 510
 500
 50 Val Ser Ile Asp Thr Val Thr Ser Tyr Arg Phe Leu Ile Gln Asp Val 525
 515
 Val Arg Gly Asp Asp Asn Lys Cys Trp Ser Pro Gln Thr Asp Lys Lys 540
 530
 55 Thr Val Ile Lys Asn Lys Pro Ser Glu Val Glu Asn Glu Ile Thr Asn 555
 545
 Ser Ile Asp Glu Lys Asp Leu Asp Ser Ser Asp Pro Leu Ile Lys Thr 575
 565
 60 Gly Leu Asp Asp Asp Ser Asp Cys Asp Pro Val Ser Glu Asn Ser Ile 585
 580
 65 Gly Gly Cys Ala Val Arg Gln Gln Leu Val Ile Tyr Thr Ala Gly Thr 605
 595
 600

WO 9507706

PCT/US94/10151

Leu His Ile Val Val Val Val Ser Ile Val Gly Leu Phe Ser Trp
610 615 620
5 Leu Tyr Ser Gly Leu Ser Val Phe Ala Lys Phe His Ser Asp Ser Gln
625 630 635 640
Tyr Pro Glu Ala Pro Phe Ile Glu Gln His Asn His Leu Glu Arg Leu
645 650 655
10 Ser Ala Asn Gln Thr Gly Tyr Leu Thr Pro Arg Ala Asn Lys Ala Val
660 665 670
15 Asn Leu Val Val Lys Val Ser Ser Thr Pro Arg Pro Lys Lys Asp
675 680 685
Asn Leu Asp Val Ser Lys Asp Leu Asn Ile Ala Ser Asp Gly Thr Leu
690 695 700
20 Gln Lys Ile Lys Lys Thr Tyr Ile
705 710

(2) INFORMATION FOR SEQ ID NO:65:

25 (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 369 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
30 (11) MOLECULE TYPE: cDNA

(15) FEATURE:

35 (A) NAME/KEY: CDS
(B) LOCATION: 1..369

(11) SEQUENCE DESCRIPTION: SEQ ID NO:65:

40 ATG ATT TAT TTA TAC ACG CCG GAT AAC GTA ATT CCA AAA CAT GGT TTA
Met Ile Tyr Leu Tyr Thr Ala Asp Asn Val Ile Pro Lys Asp Gly Leu
1 15
45 CAA GGA GCA TTT GTC GAT AAA GAC GGT ACT TAT GAC AAA GTT TAC ATT
Gln Gly Ala Phe Val Asp Lys Asp Gly Thr Tyr Asp Lys Val Tyr Ile
20 25 30
50 CTT TTC ACT GTT ACT ATC CGC TCA AAG AGA ATT GTT AAA ATT CCG TAT
Leu Phe Thr Val Thr Ile Gly Ser Lys Arg Ile Val Lys Ile Pro Tyr
35 40
55 ATA GCA CAA ATG TGC TTA AAC GAC GAA TGT GGT CCA TCA TCA TTG TCT
Ile Ala Gln Met Cys Leu Asn Asp Glu Cys Gly Pro Ser Ser Leu Ser
50 55
60 AGT CAT AGA TCG TCG ACG TTG CTC AAA GTC GAA TTA GAA TGT GAC ATC
Ser His Arg Trp Ser Thr Leu Leu Lys Val Glu Leu Glu Cys Asp Ile
65 70 75 80
65 GAC GGA AGA AGT TAT AGT CAA ATT AAT CAT TCT AAA ACT ATA AAA CAG
Asp Gly Arg Ser Tyr Ser Gln Ile Asn His Ser Lys Thr Ile Lys Gln
85 90 95
65 ATA ATC ATA CGA TAC TAT ATG TAT TCT TTG ATA CTC CTT TTC CAA GTC
Ile Met Ile Arg Tyr Tyr Met Tyr Ser Leu Ile Val Leu Phe Gln Val
100 105 110
65 CCC ATT ATG TAC CTA TTC TAT GAA TAC CAT TA

90

3750

WO 9507706

PCT/US94/10151

Arg Ile Met Tyr Leu Phe Tyr Glu Tyr His
115 120

5 (2) INFORMATION FOR SEQ ID NO:66:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 122 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(15) SEQUENCE DESCRIPTION: SEQ ID NO:66:

15 Met Ile Tyr Leu Tyr Thr Ala Asp Asn Val Ile Pro Lys Asp Gly Leu
1 5 10 15
20 Gln Gly Ala Phe Val Asp Lys Asp Gly Thr Tyr Asp Lys Val Tyr Ile
20 25 30
Leu Phe Thr Val Thr Ile Gly Ser Lys Arg Ile Val Lys Ile Pro Tyr
35 40 45
25 Ile Ala Gln Met Cys Leu Asn Asp Glu Cys Gly Pro Ser Ser Leu Ser
50 55 60
30 Ser His Arg Trp Ser Thr Leu Leu Lys Val Glu Leu Glu Cys Asp Ile
65 70 75 80
Asp Gly Arg Ser Tyr Ser Gln Ile Asn His Ser Lys Thr Ile Lys Gln
85 90 95
35 Ile Met Ile Arg Tyr Tyr Met Tyr Ser Leu Ile Val Leu Phe Gln Val
100 105 110
Arg Ile Met Tyr Leu Phe Tyr Glu Tyr His
115 120

91

3751

WHAT IS CLAIMED IS:

1. An isolated peptide of at least 5 amino acids comprising a unique portion of a semaphorin, and said peptide has a semaphorin binding specificity.
2. An isolated peptide according to claim 1 wherein said semaphorin comprises a human semaphorin.
3. An isolated antibody that specifically binds a peptide according to claim 1.
4. An isolated nucleic acid comprising a nucleotide sequence encoding a peptide according to claim 1 wherein said sequence is joined to a nucleotide not naturally joined to said sequence and said sequence is other than that of the A39 ORF of vaccinia virus.

15

5. A cell comprising a nucleic acid according to claim 3.
6. A transgenic rodent comprising a nucleic acid according to claim 7 wherein said nucleic acid is xenogeneic to said rodent.
7. A process for the production of a recombinant unique portion of a semaphorin comprising culturing the cell of Claim 4 under conditions suitable for the expression of said peptide, and recovering said peptide.

20

8. A method of identifying a pharmacological agent useful in the diagnosis or treatment of disease associated with the binding of a semaphorin to a semaphorin receptor, said method comprising the steps of:
 - contacting a panel of prospective agents with a peptide according to claim 1;
 - measuring the binding of a plurality of said prospective agents to said peptide;
 - identifying from said plurality a pharmacological agent which specifically binds said peptide;

92

3 7 5 2

wherein said pharmacological agent is useful in the diagnosis or treatment of disease associated with the binding of a semaphorin to a cellular receptor.

9. A method of diagnosing a patient for a predisposition to neurological disease associated with a genetic locus, said method comprising the steps of:

- isolating somatic cells from a patient;

- isolating genomic DNA from said somatic cells;

- contacting said genomic DNA with a probe comprising a DNA

- sequence encoding a peptide according to claim 1 under conditions wherein said probe hybridizes to homologous DNA;

- identifying a region of said genomic DNA which hybridizes with said probe;

- wherein the presence, absence or sequence of said region correlates with a predisposition to a neurological disease.

15

10. A method of treating a patient with neurological injury or disease or a pathological viral infection, said method comprising the steps of:

- administering to a patient a therapeutically effective dosage of a

- pharmaceutical composition comprising a pharmaceutically acceptable carrier and a peptide according to claim 1;

- wherein said peptide modulates neural cell growth cone function or viral pathogenicity in said patient.

11. An isolated polypeptide comprising an amino acid sequence substantially

- similar to that of a semaphorin, and said polypeptide has a semaphorin binding specificity.

25

12. An isolated peptide of at least about 5 amino acids comprising a unique portion of a semaphorin receptor, and said peptide has a semaphorin receptor binding specificity.

30

13. An isolated antibody that specifically binds a peptide according to claim 12.

93

3 7 5 3

14. An isolated nucleic acid comprising a nucleotide sequence encoding a peptide according to claim 12 wherein said sequence is joined to a nucleotide not naturally joined to said sequence.

15. A cell comprising a nucleic acid according to claim 14.

16. A process for the production of a recombinant unique portion of a semaphorin receptor peptide according to claim 12 comprising culturing the cell of Claim 14 under conditions suitable for the expression of said peptide, and recovering said peptide.

17. A method of identifying a pharmacological agent useful in the diagnosis or treatment of disease associated with the binding of a semaphorin to a cellular receptor, said method comprising the steps of:

15 contacting a panel of prospective agents with a peptide according to claim 12;

measuring the binding of a plurality of said prospective agents to said peptide;

15 identifying from said plurality a pharmacological agent which specifically binds said peptide;

wherein said pharmacological agent is useful in the diagnosis or treatment of disease associated with the binding of a semaphorin to a cellular receptor.

18. A method of diagnosing a patient for a predisposition to neurological disease associated with a genetic locus, said method comprising the steps of:

isolating somatic cells from a patient;

isolating genomic DNA from said somatic cells;

contacting said genomic DNA with a probe comprising a DNA

30 sequence encoding a peptide according to claim 12 under conditions wherein said probe hybridizes to homologous DNA;

identifying a region of said genomic DNA which hybridizes with said probe;

wherein the presence, absence or sequence of said region correlates with a predisposition to a neurological disease.

19. A method of treating a patient with neurological injury or disease or a pathological viral infection, said method comprising the steps of:

5 administering to a patient a therapeutically effective dosage of a pharmaceutical composition comprising a pharmaceutically acceptable carrier and a peptide according to claim 12.

10 wherein said peptide modulates neural cell growth cone function or viral pathogenicity in said patient.

20. An isolated polypeptide comprising an amino acid sequence substantially similar to that of a semaphorin receptor, and said polypeptide has a semaphorin receptor binding specificity.

INTERNATIONAL SEARCH REPORT

 International application No.
 PCT/US94/10151

A. CLASSIFICATION OF SUBJECT MATTER

 IPC(6) : A61K 38/00; C07K 5/00; C12P 21/06; C12Q 1/00; G01N 33/53
 US CL : 435/7.1, 69.1; 530/300

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/7.1, 69.1; 530/300

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

 APS, CA, BIOSIS, EMBASE, MEDLINE, DERWENT BIOTECHNOLOGY ABSTRACTS
 search terms: semaphorin, fasciclin

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, P ----- Y, P	Cell, Volume 75, issued 31 December 1993, A.L. Kolodkin et al, "The <i>semaphorin</i> Genes Encode a Family of Transmembrane and Secreted Growth Cone Guidance Molecules", pages 1389-1399, see the entire document.	1, 2, 11 ----- 7, 8
Y	Neuron, Volume 9, issued November 1992, A.L. Kolodkin et al, "Fasciclin IV: Sequence, expression and function during growth cone guidance in the grasshopper embryo", pages 831-845, see the entire document.	1, 2, 7, 8, 11
Y	Gene, Volume 93, issued 1990, T. Deng et al, "A novel expression vector for high-level synthesis and secretion of foreign proteins in <i>Escherichia coli</i> : overproduction of bovine pancreatic phospholipase A ₂ ", pages 229-234, see the entire document.	1, 2, 7, 8, 11

☒ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:	T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reasons (as specified)	&*	document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means		
P document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

10 NOVEMBER 1994

Date of mailing of the international search report

DEC 30 1994

 Name and mailing address of the ISA/US
 Commissioner of Patents and Trademarks
 Box PCT
 Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

BRUCE CAMPBELL

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

Int. application No.
PCT/US94/10151

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Science, Volume 251, issued 15 February 1991, S.P.A. Fodor et al, "Light-Directed, Spatially Addressable Parallel Chemical Synthesis", pages 767-773, see the entire document.	8

Form PCT/ISA/210 (continuation of second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/10151

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1, 2, 7, 8, 11

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1))(July 1992)*

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claims 1, 2, 7, 8 and 11, drawn to semaphorin peptides with semaphorin binding specificity, a method for producing said peptides, and a method for screening potential pharmaceuticals using said peptides.

Group II, claim 3, drawn to an antibody against the peptide of I.

Group III, claim 4, drawn to a nucleic acid encoding a peptide of I.

Group IV, claims 5 and 6, drawn to a cell and a rodent containing the nucleic acid of III.

Group V, claim 9, drawn to a diagnostic method using the nucleic acid of III.

Group VI, claim 10, drawn to a treatment method using the peptide of I.

Group VII, claims 12, 17 and 20, drawn to semaphorin peptides having semaphorin receptor binding specificity, and a method for screening potential pharmaceuticals using said peptides.

Group VIII, claim 13, drawn to an antibody against the peptide of VII.

Group IX, claim 14, drawn to a nucleic acid encoding the peptide of VII.

Group X, claims 15 and 16, drawn to a cell containing the nucleic acid of IX and a method of producing the peptide of VII.

Group XI, claim 18, drawn to a diagnostic method using the nucleic acid of IX.

Group XII, claim 19, drawn to a treatment method using the peptide of VII.

The inventions listed as Groups I-XII do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Groups I-VI are distinct from each of groups VII-XII because I-VI and VII-XII are drawn to compositions and methods containing and utilizing two different classes of peptides, those which bind semaphorin and those which bind semaphorin receptor. The compositions and methods of I-VI do not require the compositions and methods of VII-XII, and the compositions and methods of VII-XII do not require the compositions and methods of I-VI.

Group II is distinct from each of I and III-VI because the antibody of II is not required for the methods and compositions of I and III-VI, and the methods and compositions of III-VI are not required to produce the antibody of II. While the peptide of I can be used to elicit production of the antibody of II, the peptide can be used for other purposes as well, such as the screening and treatment methods of I and VI.

Group III is distinct from each of Groups I and V, because they are related as product and process of use. The product of III can be used for several different processes, for example the divergent processes of I and V.

Group I is distinct from each of groups IV and V because the compositions and methods of I are not required for the compositions and methods of IV and V, and the compositions and methods of IV and V are not required for I. The peptides of I can be obtained without the cells of IV, for example by chemical synthesis.

Groups I and VI are distinct because the method of VI is not required for the compositions and methods of I, and the peptide of I can be used for other methods, such as the screening method of claim 8.

Groups III and IV are distinct because they are related as intermediate and final product. The intermediate (III) can be used for other purposes, such as the methods of I and V.

Groups III and VI are distinct because the composition of III is not required for the method of VI and the method of VI is not required for the composition of III.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/10151

Group IV is distinct from each of groups V and VI because the compositions of IV are not required for the methods of V and VI, and the methods of V and VI are not required to produce the compositions of IV.

Groups V and VI are distinct because the two methods require different procedures and starting materials to achieve divergent ends.

Group VIII is distinct from each of VII and IX-XII because the antibody of VIII is not required for the methods and compositions of VII and IX-XII, and the methods and compositions of IX-XII are not required to produce the antibody of VIII. While the peptide of VII can be used to elicit production of the antibody of VIII, the peptide can be used for other purposes as well, such as the screening and treatment methods of VII and XII.

Group IX is distinct from each of Groups X and XI, because they are related as product and process of use. The product of IX can be used for several different processes, for example the divergent processes of X and XI.

Group VII is distinct from each of groups IX and XI because the compositions and methods of VII are not required for the compositions and methods of XI and XI, and the compositions and methods of IX and XI are not required for VII.

Groups VII and X are related as product and process of making. The peptide of VII can be produced without the method of X, for example by chemical synthesis.

Groups VII and XII are distinct because the method of XII is not required for the compositions and methods of VII, and the peptide of VII can be used for other methods, such as the screening method of claim 17.

Groups IX and XII are distinct because the composition of IX is not required for the method of XII and the method of XII is not required for the composition of IX.

Group X is distinct from each of groups XI and XII because the compositions of X are not required for the methods of XI and XII, and the methods of XI and XII are not required to produce the compositions of X.

Groups XI and XII are distinct because the two methods require different procedures and starting materials to achieve divergent ends.

Accordingly the claims are not so linked by a special technical feature within the meaning of PCT Rule 13.2 so as to form a single inventive concept.